

THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of : Kenichi KAWARAI, et al.

Filed : Concurrently herewith

For : PACKET SWITCH, SCHEDULING DEVICE....

Serial No. : Concurrently herewith

March 20, 2001

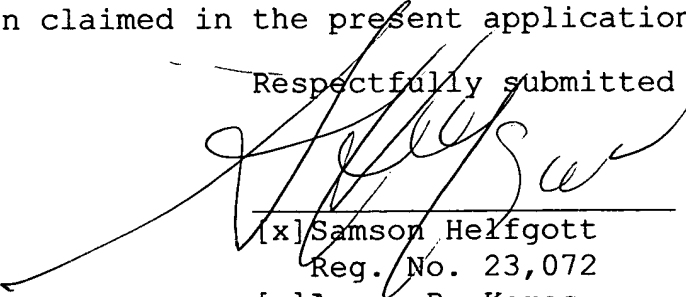
Assistant Commissioner of Patents
Washington, D.C. 20231

SUBMISSION OF PRIORITY DOCUMENT

S I R:

Attached herewith are Japanese patent application No.
2000-80383 of March 22, 2000 And 2000-124667 of April 25, 2000
whose priority has been claimed in the present application.

Respectfully submitted


[x] Samson Helfgott
Reg. No. 23,072
[] Aaron B. Karas
Reg. No. 18,923

HELFGOTT & KARAS, P.C.
60th FLOOR
EMPIRE STATE BUILDING
NEW YORK, NY 10118
DOCKET NO.: FUJA 17.344
BHU:priority

Filed Via Express Mail
Rec. No.: EL522402415US
On: March 20, 2001
By: Brendy Lynn Belony
Any fee due as a result of this paper,
not covered by an enclosed check may be
charged on Deposit Acct. No. 08-1634.

RS
2
1c973 U.S. PTO
09/813226



日 本 国 特 許 庁

PATENT OFFICE
JAPANESE GOVERNMENT

1c973 U.S. PTO
09/813226
03/20/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application:

2000年 3月22日

出 願 番 号
Application Number:

特願2000-080383

出 願 人
Applicant (s):

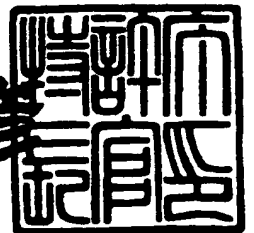
富士通株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

2000年 6月 9日

特許庁長官
Commissioner,
Patent Office

近 藤 隆 彦



【書類名】 特許願

【整理番号】 9951494

【提出日】 平成12年 3月22日

【あて先】 特許庁長官 近藤 隆彦 殿

【国際特許分類】 H04L 12/56

【発明の名称】 パケットスイッチ

【請求項の数】 15

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

【氏名】 瓦井 健一

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

【氏名】 朝永 博

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

【氏名】 松岡 直樹

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

【氏名】 加藤 次雄

【特許出願人】

【識別番号】 000005223

【氏名又は名称】 富士通株式会社

【代理人】

【識別番号】 100077517

【弁理士】

【氏名又は名称】 石田 敬

【電話番号】 03-5470-1900

【選任した代理人】

【識別番号】 100092624

【弁理士】

【氏名又は名称】 鶴田 準一

【選任した代理人】

【識別番号】 100100871

【弁理士】

【氏名又は名称】 土屋 繁

【選任した代理人】

【識別番号】 100082898

【弁理士】

【氏名又は名称】 西山 雅也

【選任した代理人】

【識別番号】 100081330

【弁理士】

【氏名又は名称】 樋口 外治

【手数料の表示】

【予納台帳番号】 036135

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9905449

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 パケットスイッチ

【特許請求の範囲】

【請求項 1】 複数の入力回線からの可変長パケットを固定長パケットに分割するパケット分割部と、

複数の入力回線のそれぞれに対して設けられ、その各々が、複数の出力回線の各々について、指定可能な Q o S (Quality of Service) クラス数よりも少数の相異なる優先度が与えられたキューを有し、対応する入力回線からの固定長パケットを、出力回線および Q o S クラスに従って該キューのいずれかに登録する複数の入力バッファ部と、

入力バッファ部のキューに登録された固定長パケットを、キューに与えられた優先度に従って単位時間内では出力回線が同じである 2 以上の固定長パケットが読み出されないように読み出すスケジューラと、

スケジューラが読み出した固定長パケットの各々を複数の出力回線のうちの指定された 1 つにルーティングするスイッチと、

複数の出力回線のそれぞれに対して設けられ、スイッチから出力される固定長パケットからの可変長パケットの組み立ておよび Q o S クラスに基づく優先制御を行なうための複数の出力バッファ部とを具備するパケットスイッチ。

【請求項 2】 前記入力バッファ部の各々は、各出力回線について第 1 の優先度のキューおよび第 2 の優先度のキューを有し、

帯域保証クラスのパケットは第 1 の優先度のキューに登録され、

ベストエフォートクラスのパケットは第 2 の優先度のキューに登録され、

最低帯域保証クラスのパケットは第 2 の優先度のキューに登録され、

第 2 の優先度のキューは、指定帯域を超えて入力された最低帯域保証クラスのパケットとベストエフォートクラスのパケットとを廃棄するための第 1 の廃棄レベルと、第 1 の廃棄レベルより高く設定され指定帯域内で入力された最低帯域保証クラスのパケットを廃棄するための第 2 の廃棄レベルとを有する請求項 1 記載のパケットスイッチ。

【請求項 3】 前記出力バッファ部の各々は、

入力回線の各々について、前記入力バッファ部における各出力回線に対するキューの数と同数のキューを有し、スイッチから出力される固定長パケットを入力回線およびQoSクラスに応じていずれかのキューに格納することによって固定長パケットから可変長パケットを組み立てる可変長パケット組立てバッファと、

指定可能なQoSクラス数と同数のキューを有し、可変長パケット組立てバッファで組み立てられた可変長パケットをQoSクラスに応じて対応するキューに登録し、パケット長およびQoSクラスに応じて可変長パケットを順次読み出して対応する出力回線へ出力するQoS制御部とを含む請求項1記載のパケットスイッチ。

【請求項4】 前記出力バッファ部の各々は、

指定可能なQoSクラス数と同数のキューを有し、スイッチから出力される固定長パケットをQoSクラスに従っていずれかのキューに登録する固定長ベースQoS制御部と、

入力回線の各々について、指定可能なQoSクラス数と同数のキューを有し、固定長ベースQoS制御部から出力される固定長パケットを入力回線およびQoSクラスに従って対応するキューに登録して可変長パケットを組み立て、対応する出力回線へ送出する可変長パケット組立てバッファとを含む請求項1記載のパケットスイッチ。

【請求項5】 前記スケジューラは、単位時間内で、前記複数の入力バッファ部内のすべてのキューを、キューに与えられた優先度の順、入力回線の所定の順番、および出力回線の所定の順番で調べて、各出力回線へ出力する固定長パケットを1つずつ選択するものであり、

同一の優先度、同一の入力回線のキューの中から固定長パケットを送出する出力回線を選択するにあたり、可変長パケットを構成する複数の固定長パケットの一部が送出済である出力回線に対応するキュー内の固定長パケットを他に優先して選択する請求項1記載のパケットスイッチ。

【請求項6】 前記出力バッファ部の可変長ベースQoS制御部は、

可変長パケットがいずれかのキューの先頭に達したとき、そのパケット長に応じた値を対応するトークンに加算する手段と、

出力回線へのパケット送出が可能なとき、可変長パケットが存在しているキューの中で、対応するトークンが最小もしくは0であるものからの可変長パケットの読み出しを開始する手段と、

単位時間が経過するごとに、可変長パケットが存在しているキューのQoSクラスに与えられたウェイトの総和に対する、可変長パケットが存在しているそれぞれのキューのQoSクラスに与えられたウェイトの比を、対応するトークンから減算する手段とを含む請求項3記載のパケットスイッチ。

【請求項7】 前記出力バッファ部の可変長ベースQoS制御部は、

可変長パケットがいずれかのキューの先頭に達したとき、そのパケット長と可変長パケットが存在しているキューのQoSクラスに与えられたウェイトの総和との積に応じた値を対応するトークンに加算する手段と、

出力回線へのパケット送出が可能なとき、可変長パケットが存在しているキューの中で、対応するトークンが最小もしくは0であるものから可変長パケットの読み出しを開始する手段と、

単位時間が経過するごとに、可変長パケットが存在しているそれぞれのキューのQoSクラスに与えられたウェイトを、対応するトークンから減算する手段とを含む請求項3記載のパケットスイッチ。

【請求項8】 前記出力バッファ部の可変長ベースQoS制御部は、

可変長パケットがいずれかのキューの先頭に達したとき、可変長パケットが存在しているキューのQoSクラスに与えられたウェイトの総和に応じた値を単位時間が経過するごとに加算し、これをそのパケット長に対応する回数だけ繰り返す手段と、

出力回線へのパケット送出が可能なとき、可変長パケットが存在しているキューの中で、対応するトークンが最小もしくは0であるものから可変長パケットの読み出しを開始する手段と、

単位時間が経過するごとに、可変長パケットが存在しているそれぞれのキューのQoSクラスに与えられたウェイトを、対応するトークンから減算する手段とを含む請求項3記載のパケットスイッチ。

【請求項9】 前記出力バッファ部の可変長ベースQoS制御部は、

帯域保証クラスの変長パケットおよび指定帯域内で入力された最低帯域保証クラスの変長パケットの固定長パケット数をQoSクラスごとにカウントする第1のカウントと、

ベストエフォートクラスの変長パケットおよび指定帯域を超えて入力された最低帯域保証クラスの変長パケットの固定長パケット数をカウントする第2のカウントと、

第1のカウントについて0でないカウント値を有するQoSクラスおよび第2のカウントについて0でないカウント値を有するQoSクラスを対象としてパケット長一定として優先して出力すべきパケットのQoSクラスを決定して通知する優先制御部と、

優先制御部からの通知をQoSクラスごとにカウントし、カウント値が対応するキューの先頭の変長パケットのパケット長に相当する値に達したQoSクラスについて、変長パケットを読み出して出力回線へ送出する変長パケット管理部とを含む請求項3記載のパケットスイッチ。

【請求項10】 バッファ内に格納されているパケットの量のその閾値との差分にバッファの先頭に存在する変長パケットのパケット長を乗じてその結果を出力する演算回路と、

カウントと、

演算回路の演算結果をカウントに加算する加算回路と、

カウントの値が所定値を超えているときバッファの先頭に存在する変長パケットを一度に廃棄し、前記パケット長に所定値を乗じた値をカウントから減算する制御回路とを具備する変長パケットの廃棄制御回路。

【請求項11】 バッファ内に格納されているパケットの量のその閾値との差分を出力する演算回路と、

カウントと、

演算回路の演算結果をカウントに加算する加算回路と、

カウントの値が所定値を超えているときバッファの先頭に存在する固定長パケットを廃棄し、所定値をカウントから減算する制御回路とを具備する変長パケットの廃棄制御回路。

【請求項 1 2】 バッファに格納されている、マルチキャスト制御すべき可変長パケットのアドレスを登録するマルチキャストキューと、

複数の出力回線のそれぞれに対して設けられた複数の出力バッファキューと、
バッファに格納されている可変長パケットのアドレスを管理するアドレス管理テーブルと、

マルチキャスト制御すべき可変長パケットのアドレスがマルチキャストキューに登録されたとき、該可変長パケットのアドレスを含むレコードをその出力先の数だけアドレス管理テーブルに格納し、その出力先に対応する出力バッファキューに、該レコードのアドレスをそれぞれ登録する制御手段とを具備するマルチキャスト制御回路。

【請求項 1 3】 Q o S クラスを有する可変長パケットをスイッチングするパケットスイッチにおいて、

複数の Q o S クラスを簡易な優先度クラスにマッピングする手段と、
簡易な優先度クラスに基づいて、前記可変長パケットの読み出し制御を行う手段を備えたことを特徴とするパケットスイッチ。

【請求項 1 4】 Q o S クラスを有する I P パケットを固定長パケットへ変換し、固定長パケット単位でスイッチングするパケットスイッチにおいて、

I P パケットの優先度クラスを、I P パケットに割り当てられる Q o S クラスの数より少ない数の簡易な優先度クラスにマッピングする手段と、

簡易な優先度クラスに基づいて、前記固定長パケットの読み出し制御を行う手段を備えたことを特徴とするパケットスイッチ。

【請求項 1 5】 Q o S クラスを有する I P パケットを固定長パケットへ変換し、固定長パケット単位でスイッチングするパケットスイッチにおいて、

I P パケットの優先度クラスを、I P パケットに割り当てられる Q o S クラスの数より少ない数の簡易な優先度クラスにマッピングする手段と、

簡易な優先度クラスに基づいて、前記固定長パケットをスイッチへ読み出し制御を行う手段と、

スイッチング後の前記固定長パケットを、I P パケットに割り当てられる Q o S クラスに基づいて読み出し制御する手段を備えたことを特徴とするパケットス

イッチ。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、I P (Internet Protocol) パケットのような可変長パケット（またはフレーム）のルーチングを行なうパケットスイッチ、特に、可変長パケットに対するQ o S (Quality of Service) 制御、廃棄制御、マルチキャスト制御といった複雑な制御を実現する大容量パケットスイッチに関する。

【0002】

【従来技術】

近年、インターネットの急激な普及により、通信トラヒック需要が急増しており、T b p s (テラビット/秒) クラス以上の高速・大容量のルータやスイッチの実現が望まれている。

従来、I P パケットのような可変長パケットを処理するルータ装置において、パケットフォワーディング処理をハードウェア化することにより、高速化を図るアプローチが行われているが、Q o S 制御、廃棄制御、マルチキャスト制御といった複雑な処理については、ソフトウェア処理、またはハードウェアで行う場合にも低速での処理に限られていた。またA T M (Asynchronous Transfer Mode) 交換機においては、基本的に固定長パケット毎にQ o S 制御、廃棄制御、マルチキャスト制御が行われていた。

【0003】

前者のルータ装置のようにソフトウェアベースの方式では、複雑な処理を実装することができるが、回線速度の高速化にともない、回線速度と同等の速度でソフトウェアの処理を実現することは困難となっており、より高速・大容量の処理は困難となっていた。

後者のA T M 交換機では、固定長パケット毎にQ o S 制御、廃棄制御、マルチキャスト制御を行うため、I P パケットのような可変長パケット毎にQ o S 制御、廃棄制御、マルチキャスト制御する機能を持たない。またG F R (Guaranteed Frame Rate) というA T M においてA A L 5 の可変長パケット識別を利用して

、可変長パケットレベルのQoS保証、廃棄制御を行うという試みも行われているが、この方式でも、パケットのスイッチング、転送単位はセルと呼ばれる固定長パケット単位であり、ネットワークを構成する全ノードをATM交換機で実現する必要があり、IPパケット単位でスイッチング、転送処理を行うパケットスイッチとして適用することはできない。

【0004】

次に、QoS制御に関しては、主にベストエフォートクラスの可変長パケット間での公平な帯域分配方法として、WFQ (Weighted Fair Queueing) が知られている (A.K.Parekh, R.G.Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Service Networks: The Single-Node Case", IEEE/ACM trans. On Networking, Vol 1, No.3, pp.344-357, Jun. 1993; A. K.Parekh, R.G.Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Service Networks: The Multiple-Node Case", IEEE/ACM trans. On Networking, Vol 2, No.2, pp.137-150, Jun. 1993)。

【0005】

WFQは、キューイングされているコネクションに設定されているウェイトに比例して余剰帯域を分配するGPS (Generalized Processor sharing) を基本にしている。GPSでは、アクティブな(キューを使用している)任意のコネクション(i)(j)について常に式(1)が成立するように読出しデータ量を分配する帯域割り当て方式である。

【0006】

$$S(i, \tau, t) / S(j, \tau, t) \geq \Phi(i) / \Phi(j) \quad \dots (1)$$

ここで、 $\Phi(N)$ はコネクションNに対するウェイト、 $S(N, \tau, t)$ はコネクションNに対して時間 $[\tau, t]$ の間にサービスされるデータ量である。また式(1)で等号が成り立つのはコネクション(i)と(j)のデータがともにキューイングされているときである。例えば、各コネクション#0, #1, #2, #3にそれぞれ、ウェイトを0.1:0.2:0.4:0.3の割合で与えておき、もし仮にコネクション#3のみノンアクティブ(パケットがキューイングされていない)で、それ以外のコネクションがアクティブであるとき、コネクシ

ョン#3を除くコネクション#0～#2のコネクションで、各々読出し帯域が0.1:0.2:0.4になるように帯域を分配する。

【0007】

GPSは基本的に無限小のビット単位での帯域分配を意味するが、現実には分配単位が可変長パケット（フレーム）毎であるため、WFQではGPSをフレームレベルの公平な帯域分配に拡張している。

WFQでは、フレーム到着時にフロー*i*の*k*番目のパケットが時刻*t* (*i*, *k*)に到着し、そのフレーム長を*L* (*i*, *k*)としたときの終了予定時刻*F* (*i*, *k*)を式(2)で計算し、*F*の小さい順から読み出す。

【0008】

$$F(i, k) = \max \{ F(i, k-1), t(i, k) \} + L(i, k) / r(i, k) \quad \dots (2)$$

但し、 $r(i) = \Phi(i) / \sum \Phi \times R$ 、ここで $\sum \Phi$ はキューイングされているコネクションのウェイトの合計であり、*R*は出力リンクのレートである。

ここで、*r* (*i*, *k*)はパケットがキューから読み出される度に変化するため、全コネクションについて*F* (*i*, *k*)の再計算が必要となり、また、キューイングされている全パケットについて、終了予定時刻*F* (*i*, *k*)を保持しなければならないため、ハード規模が大きくなるという問題があった。そのため実時間の代わりにVirtual Time（キューイングされているコネクション数に応じて刻みの変化する時間カウンタ）を式(3)のように定義し、*F* (*i*, *k*)を式(4)のように再定義し、Virtual Timeを $\sum \Phi$ と*R*が変化するたびに再計算し、*F* (*i*, *k*)については、パケット到着時に1回計算するのみで実現する方式が提案されている。

【0009】

$$Vt = V(\tau) + (R / \sum \Phi) (t - \tau) \quad \dots (3)$$

$$F(i, k) = \max \{ F(i, k-1), Vt(i, k) \} + L(i, k) / \Phi(i) \quad \dots (4)$$

しかしながら、Virtual timeを用いると、演算量の削減が図れるが、実時間で処理する必要のある転送レートを規定した帯域保証型のQoSスケジューラとの

間の優先度制御が困難となるという問題が生じていた。

【0010】

廃棄制御に関しては、アダプティブな廃棄方式としてRED (Random Early Detection) が知られている (Internet Society RFC2309)。これは、輻輳状態が同期してしまうことを防ぐため、輻輳状態になったらバッファが溢れる前に輻輳状態に応じた確率でランダムに廃棄を行う方式である。これにより、TCPのような輻輳制御機能を持つプロトコルでは早めに送信レートを落とすことが出来、かつ全端末の動作が同期しないので安定したスループットが出るという特長を持つ。

【0011】

しかしながら、REDは、廃棄をランダムに行う必要があるためアルゴリズムが複雑でソフトウェアにより実装されており、高速化には向いていなかった。また、基本的に廃棄はフレーム (可変長パケット) 到着時に行うため、その時点でバッファに溜まっているデータが全て掃けた後でないと廃棄が端末に通知されないという問題があった。

【0012】

マルチキャスト制御に関しては、バッファアドレス管理方式として、従来では固定長パケットに対するパケット by パケットの動作のみが実現されており、複数の固定長パケットの連続からなる可変長パケットを意識した制御は実現されていない。

【0013】

【発明が解決しようとする課題】

したがって本発明の目的は、可変長パケットに対して、高速かつ少ないハードウェア規模でQoS制御、廃棄制御、マルチキャスト制御といった複雑な制御を実現する大容量パケットスイッチを提供することにある。

【0014】

【課題を解決するための手段】

本発明によれば、複数の入力回線からの可変長パケットを固定長パケットに分割するパケット分割部と、複数の入力回線のそれぞれに対して設けられ、その各

々が、複数の出力回線の各々について、指定可能なQoS (Quality of Service) クラス数よりも少数の相異なる優先度が与えられたキューを有し、対応する入力回線からの固定長パケットを、出力回線およびQoSクラスに従って該キューのいずれかに格納する複数の入力バッファ部と、入力バッファ部のキューに格納された固定長パケットを、キューに与えられた優先度に従って単位時間内では出力回線が同じである2以上の固定長パケットが読み出されないように読み出すスケジューラと、スケジューラが読み出した固定長パケットの各々を複数の出力回線のうちの指定された1つにルーティングするスイッチと、複数の出力回線のそれぞれに対して設けられ、スイッチから出力される固定長パケットからの可変長パケットの組み立ておよびQoSクラスに基づく優先制御を行なうための複数の出力バッファ部とを具備するパケットスイッチが提供される。

【0015】

本発明によれば、バッファ内に格納されているパケットの量のその閾値との差分にバッファの先頭に存在する可変長パケットのパケット長を乗じてその結果を出力する演算回路と、カウンタと、演算回路の演算結果をカウンタに加算する加算回路と、カウンタの値が所定値を超えているときバッファの先頭に存在する可変長パケットを一度に廃棄し、前記パケット長に所定値を乗じた値をカウンタから減算する制御回路とを具備する可変長パケットの廃棄制御回路もまた提供される。

【0016】

本発明によれば、バッファ内に格納されているパケットの量のその閾値との差分を出力する演算回路と、カウンタと、演算回路の演算結果をカウンタに加算する加算回路と、カウンタの値が所定値を超えているときバッファの先頭に存在する固定長パケットを廃棄し、所定値をカウンタから減算する制御回路とを具備する可変長パケットの廃棄制御回路もまた提供される。

【0017】

本発明によれば、バッファに格納されている、マルチキャスト制御すべき可変長パケットのアドレスを格納するマルチキャストキューと、複数の出力回線のそれぞれに対して設けられた複数の出力バッファキューと、バッファに格納されて

いる可変長パケットのアドレスを管理するアドレス管理テーブルと、マルチキャスト制御すべき可変長パケットのアドレスがマルチキャストキューに格納されたとき、該可変長パケットのアドレスを含むレコードをその出力先の数だけアドレス管理テーブルに格納し、その出力先に対応する出力バッファキューに、該レコードのアドレスをそれぞれ格納する制御手段とを具備するマルチキャスト制御回路もまた提供される。

【 0 0 1 8 】

本発明によれば、QoSクラスを有する可変長パケットをスイッチングするパケットスイッチにおいて、複数のQoSクラスを簡易な優先度クラスにマッピングする手段と、簡易な優先度クラスに基づいて、前記可変長パケットの読み出し制御を行う手段を備えたことを特徴とするパケットスイッチもまた提供される。

本発明によれば、QoSクラスを有するIPパケットを固定長パケットへ変換し、固定長パケット単位でスイッチングするパケットスイッチにおいて、IPパケットの優先度クラスを、IPパケットに割り当てられるQoSクラスの数より少ない数の簡易な優先度クラスにマッピングする手段と、簡易な優先度クラスに基づいて、前記固定長パケットの読み出し制御を行う手段を備えたことを特徴とするパケットスイッチもまた提供される。

【 0 0 1 9 】

本発明によれば、QoSクラスを有するIPパケットを固定長パケットへ変換し、固定長パケット単位でスイッチングするパケットスイッチにおいて、IPパケットの優先度クラスを、IPパケットに割り当てられるQoSクラスの数より少ない数の簡易な優先度クラスにマッピングする手段と、簡易な優先度クラスに基づいて、前記固定長パケットをスイッチへ読み出し制御を行う手段と、スイッチング後の前記固定長パケットを、IPパケットに割り当てられるQoSクラスに基づいて読み出し制御する手段を備えたことを特徴とするパケットスイッチもまた提供される。

【 0 0 2 0 】

【発明の実施の形態】

図1に本発明の一実施形態に係るパケットスイッチを概念的に示す。入力回線

から入力された可変長パケットは、パケット分割部10において固定長パケットに分割され入力バッファ部12にバッファリングされる。入力バッファ部12は、N本の入力回線数分あり、各入力バッファは論理的にM本の出力回線毎、およびQoS毎（帯域保証クラス11、ベストエフォートクラス17）に分かれている。各入力回線より入力バッファ部12へ入力されたパケットは各々対応する出力回線毎、QoS毎にバッファリングされる。この際、QoSクラスを予め契約や呼受け付け制御により入力トラヒック量の予想可能な帯域保証クラスと、契約や呼受け付け制御がなく入力トラヒック量の予想ができないベストエフォートクラスの2クラスのいずれかに縮退して割り当てる。

【0021】

入力バッファ部12にバッファリングされた各パケットは、回線間スケジューラ14により、スイッチ部16への送出タイミングが調整される。この際、回線間スケジューラ14では、はじめに出力回線毎にバッファリングされたパケットのうち帯域保証クラスのパケットについて、複数の入力回線で同一の出力回線を選択しないように出力回線を選択する、競合制御を行う（図1：13）。次にベストエフォートクラスのパケットについて、このとき選択されなかった出力回線について、同様に複数の入力回線で同一の出力回線を選択しないように競合制御を行う（図1：15）。以上のような回線間スケジューリング動作を単位時間内、例えば、1つの出力回線で固定長パケット1個に相当するデータを送出するに要する時間内で実施する。これにより、各入力回線間の論理出力バッファ毎、QoSクラス毎の競合制御のもとで各入力バッファから出力されるパケットが選択され、スイッチ部16へ転送される。転送されたパケットはスイッチ部16でパケットに付与されたあて先情報を元に各出力バッファ部18へスイッチングされる。

【0022】

出力バッファ部18に入力した各パケットは、実際のQoSクラス毎（通常3クラス以上）にバッファリングされる。出力バッファ部18では、各出力回線内の各QoSクラス毎のQoS制御（設定レートや優先順位に応じた読み出し順序制御）を行い、各出力回線にパケットを出力する。

ここでは、入力バッファ部12でのQoSクラス数を帯域保証クラスとベストエフォートクラスの2クラスに縮退させている。これにより、多数のキュー間（入力回線数×出力回線数×QoSクラス数（3以上））での競合制御を入力回線数×出力回線数×2に削減できる。

【0023】

一般にQoS制御としては、単純な優先制御だけでなく、読み出しレートなどの複雑な読み出し順序の制御が必要であるが、QoSクラスを帯域保証クラスとベストエフォートクラスに縮退させることにより、帯域保証クラスをスケジューリングしてからベストエフォートクラスをスケジューリングすることによりスケジューリング処理を簡単化している。

【0024】

さらに、スイッチの内部動作速度を入力バッファの到着速度より高め、スイッチング後の出力バッファ18において、実際のQoSクラス毎のQoS制御を行うことにより、パケットスイッチトータルとしてのQoS制御ポイントを出力バッファ部のQoS制御に集約できる。この際、出力バッファ部では、各出力回線毎のQoSクラス制御のみ行えば良いことから、スイッチ全体のQoS制御が必要な入力バッファ部で行う場合に比べ、ハード量は少ない。

【0025】

図1のパケットスイッチで採用されている、出力回線毎の論理キューを持つ入力バッファ型スイッチでは、ほぼ100%のスループットが実現されるので、予めトラヒック量が予想できるクラスに対しては、入力バッファ側におけるパケットの廃棄率や遅延時間を出力バッファ側におけるQoS制御によって生じるパケット廃棄率や遅延時間に比べて非常に小さくできる。スイッチ内部の動作速度を高めることにより、これらはさらに小さくすることができる。したがって、入力側で複数のQoSクラスを2クラスにマッピングしても実質的に問題ない。

【0026】

図2および図3にそれぞれ本発明における可変長パケットフォーマットの一例、および可変長パケットを複数の固定長パケットに分離するときの固定長パケットのフォーマットの一例を示す。また、表1および表2にはそれぞれ、図2およ

び図3における各フィールドの説明を示す。ここでは、可変長パケットとしてIPv4フォーマット（Internet Society RFC791）を用いている。但し、各パケットのフローの識別が可能なアドレス、およびQoSクラス値が含まれているパケットであれば、どのようなフォーマットでも良い。ここでQoSクラス値は、パケット中に記述されたQoSクラス値ではなく、フローやコネクションに対して予め設定されたテーブルからQoSクラスを変換して求めても良い。固定長パケットも同様に装置内部ローカルなパケットであるため、必要な情報要素が含まれていれば、フォーマットはどのようなフォーマットでも良い。

表1

Vers	バージョン番号
IHL	ヘッダ長
Type of Service	優先度
Total Length	IPv4ペイロード長
Identification	フラグメントID
Flags	フラグ
Fragment Offset	フラグメントオフセット
Time to Live	ホップ期限
Protocol	プロトコル
Header Checksum	ヘッダチェックサム
Source Address	送信元アドレス
Destination Address	宛先アドレス
Payload	ペイロード

表2

EN	空きセル識別子（0：空きセル、1：有効セル）
C	マルチキャスト識別子
QCP	Quality of Service：品質クラス識別子
CID	コネクション識別子
PKL	ペイロード長
F	タギングフラグ（0：高優先クラス、1：低優先クラス）

F R I フレームタイプ (0 0 : 中間、0 1 : 先頭、1 0 : 最終、1 1 : 先頭 & 最終)

図4に図1のスイッチ部16の入力側におけるスケジューリングを実現する具体的な構成を示す。図4の入力バッファ部のスケジューラ構成において、入力されたフレームは、ポリシング制御部20において、予め設定されたレート以上のものについてはタギングされる。タギング情報はフレーム毎に付与され、別線、または図2の可変長パケットフォーマットに内部ヘッダを追加してそこにビットを付与することにより後段に転送される。パケット分割部10では入力された可変長パケットを一定長のセグメントに分割する。ヘッダ付与部22では、ヘッダ変換テーブル24の情報をもとに各セグメントに各種ヘッダ情報を付与して図3の固定長パケットを組み上げる。

【0027】

ここで、QoSクラスに関しては、予めコネクション毎、またはフロー毎に設定されたQoSクラス（例えば4クラス）のテーブルを参照しQCPビットに0～3の値を付与するとともに、これらのQoSクラスをさらに帯域保証クラスとベストエフォートクラスの2クラスにマッピングする。この2クラスにマッピングされたQoSクラスを縮退QoSクラスと呼ぶことにする。入力バッファ部12、およびスケジューラ部25ではこの縮退QoSクラスと出力回線番号情報に基づき、それぞれパケットのキューイング、パケット読み出しのスケジューリングを行う。

【0028】

ここで、スケジューラ部25では、後段のマトリックススイッチ部へ入力される全パケットについて、効率良くしかも同一の出力回線を持つパケットが同時に転送されないようにスケジューリングを行う。さらにこの競合制御時に縮退QoSクラスに応じて、初めに帯域保証クラスについてスケジューリングを行い、次に帯域保証クラスのスケジューリングで選択されなかった入力回線と出力回線についてベストエフォートクラスについてスケジューリングを行う。

【0029】

図5は、スケジューラの正順、逆順を考慮したパイプラインの8回線について

帯域保証クラス、ベストエフォートクラスをスケジューリングした場合のパイプラインのシーケンスである。

図 5 に示した例では、パイプラインのシーケンス順序を入れ替えることにより、1 周期に正順、逆順を考慮したパイプラインと同じパターンが同数現れるようになっており、入力回線間の公平性を保証している。さらに同一時間に同一入力回線に対する処理の競合を避けることになり、同じ処理をしなければならないハード処理量を削減している。

【 0 0 3 0 】

尚、スケジューリングアルゴリズムについては、帯域保証クラスの方がベストエフォートクラスより優先度が高くなれば、他のアルゴリズムを用いても良い。

以上のようなスケジューリング処理により決定されたパケットは入力バッファ部からよみだされ、マトリックススイッチにおいて、出回線情報にもとづいてスイッチングされる。このとき、マトリックススイッチのスイッチング速度を各回線のデータレートの総和より大きく（内部速度アップという）することにより、入力バッファ部の廃棄遅延特性の向上を図り、入力バッファ部における競合によりスイッチトータルとしての品質が劣化するのを抑制してもよい。

【 0 0 3 1 】

スイッチングされた固定長パケットは一旦フレーム組立てバッファにキューイングされる（図 1 : 1 8）。この際、後に説明するようにフレーム組立てバッファ部において、フレームへ戻すかまたは先に固定長パケットのままで、Q o S 制御を行ってからフレーム組立てを行っても良い。

以上のような一連の処理により、入力バッファ側では、縮退した 2 クラスのみ簡易な Q o S 制御を行いながら、出力バッファ側で 3 クラス以上の Q o S 制御を行うことにより、スイッチトータルとして、3 クラス以上の Q o S 制御を可能としている。

【 0 0 3 2 】

上記は、入力フレームを一旦固定長パケットに分割して行っているが、スケジューラ部において、フレームの連続性を崩さない選択を行うことにより、実質的にフレームのままでスイッチング処理を行なうようにしても良い。

図 6 は、1 Q o S クラス内に帯域保証が必要な最低保証レート分と帯域保証が不要だがトラフィック量が予測できないベストエフォート分を合わせ持つ、最低帯域保証クラスを収容する手法の一例を示す。最低帯域保証クラスを収容する際、最低帯域保証クラスの packets もベストエフォートクラスのキューにバッファリングする。但し、キューの前段のポリシング制御部 2 6 において、最低帯域保証クラスに対して、最低保証レートを超過して入力される packets に対してはタギングを行っておき、タギングなしの packets の廃棄閾値 θ_3 に対して、タギングありの packets とベストエフォートクラスの packets に対する、使用可能なキューのしきい値 θ_2 を低く設定する。

【 0 0 3 3 】

これにより最低帯域保証クラスを含む場合についても 2 クラスにマッピングすることが可能になる。この際、最低帯域保証クラスの最低保証レート内で入力される packets の廃棄閾値 θ_3 を帯域保証クラスの packets の廃棄閾値 θ_1 と等しく設定して、帯域保証クラスと同等の品質保証を可能としても良い。また最低帯域保証クラスの最低保証レートを超過して入力される packets については、同じベストエフォートクラスにマッピングされる。

【 0 0 3 4 】

図 7 にフローチャートを示す。入力バッファ部において、到着 packets が最低帯域保証クラスのときに、ベストエフォートクラスとしてキューイングしておき、その際図 7 のようにタギング情報に基づいた廃棄制御を行うことにより、扱うクラス数を増やさずに帯域保証クラスの品質保持を可能としている。

図 8 および 9 に出力バッファ部 1 8 の構成の 2 つの例を示す。出力バッファ部 1 8 には複数の入力回線からの packets が混在した状態で到着する。この混在した状態を packets インタリーブまたは単にインタリーブという。従って、出力バッファ部 1 8 でこのインタリーブされた複数の固定長 packets を組立て、元の可変長 packets に戻してから、出力回線より packets を送出する必要がある。さらに Q o S 保証するためには、予め設定されたレートや優先度に応じた順序で packets が送出されなければならない。

【 0 0 3 5 】

図 8 の例では、出力バッファ部にインタリーブして到着した固定長パケットをフレーム組立てバッファ 2 8 に一旦バッファリングし、元の可変長パケットに戻す。この際インタリーブして到着したパケットは、複数の入力回線のものからのものが混在しており、さらに同一入力回線内でも入力回線側で Q o S クラス制御をした場合には、複数のクラスからのパケットが混在する。前述の例ではこの Q o S クラス数は 2 クラスとなる。従って、入力回線数 (N) × 入力回線側の Q o S クラス数 (ex. 2) 個のキューを用意し、インタリーブして到着した固定長パケットを一旦バッファリングし、元の可変長パケットを組立てる。

【 0 0 3 6 】

可変長パケットが組立て終わったら、Q o S クラス (ex. 4 クラス) に応じて、可変長ベース Q o S 制御部 3 0 に転送し、可変長ベース Q o S 制御部 3 0 において、パケット長を考慮した Q o S 制御を行うことにより、予め設定されたレートや優先度に応じた順序でパケットを出力回線より転送する。

図 9 の例では、出力バッファ部にインタリーブして到着した固定長パケットのままで、固定長ベース Q o S 制御部 3 2 においてバッファリングし、予め設定されたレートや優先度に応じた順序でパケットを後段のフレーム組立て部に転送する。この場合の Q o S 制御はパケット長が固定であるので容易である。

【 0 0 3 7 】

後段の可変長パケット組立てバッファ 3 4 には入力回線毎および、出側での Q o S クラス毎にパケットがインタリーブされた状態で到着するので、入力回線数 (N) × 出側の Q o S クラス数 (ex. 4) 個のキューを用意し、インタリーブして到着した固定長パケットを一旦バッファリングし、元の可変長パケットを組立て、パケットが組みあがった順番に出力回線より可変長パケットを転送する。

【 0 0 3 8 】

図 1 0 は図 8 をより詳しくした図である。可変長パケット組立てバッファ 2 8 に到着するパケットは、異なる入力回線、異なる縮退 Q o S クラスのものがインタリーブして到着するので、可変長パケット組立てバッファ 2 8 では、入力回線毎、縮退 Q o S クラス毎にキューイングする。可変長パケット管理部 3 6 では、各パケットに付与された可変長パケットの先頭、末尾の情報より 1 可変長パケッ

ト分のパケットを監視し、1可変長パケット分が組みあがったら、可変長パケット送信通知をだし、1可変長パケット分連続的に後段へ転送する。この際、実際のパケット内の情報を転送するのではなく、バッファに書きこまれたアドレス、およびアドレスチェーン情報のみをコピーしてもよい。さらに後段のQoS制御部30では、QoSクラス数分のバッファを用意しておき、到着パケットをQoSクラス（ex. 4クラス）毎にキューイングする。QoSスケジューリング部38では、キューへ書きこまれた、パケットのQoSクラスやフレーム長を監視し、パケット長や予め設定されたレートやウェイト、優先度に応じて読み出し順序の競合制御を行い、読み出し可能となったQoSクラスに対して出力回線への送出を指示する。

【0039】

図11は出力回線がさらに複数の低速回線へDMUX（分離）される場合を示す。フレーム組立てバッファ28の処理は同じであるが、QoS制御部30において、QoSクラス毎のキューを論理出方路（DMUX）分保持しており、QoSスケジューリング部38において、各DMUX中の各QoSクラスキューからの読み出しがDMUXの帯域を超えないように制御する。ここで、各DMUXへのタイムスロットを割り当て、タイムスロットに対応するDMUXキュー内の各QoSクラスについてQoSスケジューリングを行う。

【0040】

図12では図9の構成において、図11と同様に複数の低速回線へDMUX（分離）されるとともに、QoS制御部と可変長パケット組立てバッファの論理的な配備位置が逆になっている。このため、QoS制御部32では、QoSクラス毎、DMUX毎にキューイングし、固定長パケットのまま各DMUX中の各QoSクラスキューからの読み出しがDMUXの帯域を超えないように行なわれてQoSスケジューリングを行う。また可変長パケット組立て部34には、入力回線毎、QoSクラス毎、DMUX毎にキューイングしておき、1可変長パケット分組みあがった時点で出力回線へ1可変長パケット分連続的に出力する。

【0041】

出力バッファ部18での処理を考慮すると、スイッチ部16の各出力回線から

出力される固定長パケットの並びが、複数の入力回線からのパケットが混在した状態であるとしても、同じ可変長パケットから生じた固定長パケットができるだけ連続するように入力側でスケジューリングすることが望ましい。図 1 3 は、この点を考慮した回線間スケジューラ 1 4 の構成を示す。スケジューラ部の可変長パケット管理部 4 0 は、各論理出力方路毎の可変長パケット送出状態を管理するものであり、可変長パケットを構成する最初の固定長パケットを送出したときフラグをセットし、最終パケットを送出したときクリアすることにより、可変長パケット送出状態を識別する。また、未確定管理部 4 2 は、他の入力回線によって読み出しが確定している回線を管理するもの、要求管理部 4 4 は、入力バッファ部からの読み出し要求数を管理するもの、スケジューリング処理部 4 6 は、要求情報、未確定情報、可変長パケット送出状態をもとに送出論理出力方を決定するものである。

【 0 0 4 2 】

スケジューリング処理部 4 6 は、各論理出力回線毎に、現在可変長パケットを構成する固定長パケットを送出中であるかを判定し、スケジューリングの際に送出中の論理出力回線を優先的に選択する。これにより、1 つの可変長パケットを構成する固定長パケットがまとめて読み出されやすくなるため、出力側の可変長パケットバッファでの待ち合わせ時間を低減することが可能となり、出力側の可変長パケット組み立てバッファ量を削減することができる。

【 0 0 4 3 】

以下にスケジューリング動作の具体的な例を図 1 4 および図 1 5 を参照して説明する。ここでは、簡略化のため 3 × 3 スイッチを例にとり動作を説明する。各入力回線は、論理的に出力方路毎に分割された論理キュー（以下、VOQ : Virtual Output Queue と記す）4 7 を有している。また、可変長パケット送出状態を示す可変長パケット状態レジスタ 4 8 を各 VOQ 毎に備える。各 VOQ はさらに複数の QoS クラスごとに分割されているがここでは図示が省略されている。

【 0 0 4 4 】

上記可変長パケット状態レジスタ 4 8 は、可変長パケット送出中の VOQ を識別するためのレジスタであり、バッファからパケットを読み出したとき、そのパ

ケットのケット種別を識別し、可変長ケットエンドケットで無いとき、“1”にセットされ、可変長ケットエンドケットのときにリセットされる。スケジューラでは、本レジスタが“1”に設定されている状態を可変長ケット送出状態として扱う。読み出したケットが可変長ケットエンドケットであるか否かは、ケットヘッダのケット識別ビットを参照することにより判定可能である。

【0045】

スケジューリング処理部46は、上記可変長ケット状態レジスタ48を参照して、以下の2つのSTEPに沿って読み出しVOQを決定する。

STEP1：可変長ケット送出中のVOQのみをスケジューリング対象としてスケジューリング処理を実施。

STEP2：STEP1において読み出しVOQが確定しなかった場合に限り、可変長ケット未送出状態のVOQを対象としてスケジューリングを実施。

【0046】

まず、初期状態として図14（左側）に示すようなケットが蓄積されており、スケジューリング順序として、入力回線#0から順に入力回線#1→#2の順でスケジューリングが実行されるとする。スケジューリングの開始入力回線（図中●印）は、スケジューリング周期毎に所定の規則に従って更新される。

● T=0におけるスケジューリング動作

入力回線#0では全てのVOQが可変長ケット未送出状態であるため、上記STEP1ではどのVOQも確定されない。従って、STEP2として可変長ケット未送出回線（この例では全てのVOQ）の中からラウンドロビン方式により読み出しVOQを決定する。ここでは、VOQ#0が選択されたとする。（図中、読み出しVOQを★印で示す）入力回線#1でも同様の手順でスケジューリングが実行されるが、既に入力回線#0によってVOQ#0が選択されているため、T=0ではどのVOQも選択することが出来ない。入力回線#2では、VOQ#1が選択されたとする。

● T=0における可変長ケット状態レジスタ更新

スケジューリング処理によって決定されたVOQからケットを読み出し、読

み出したパケット種別に応じて可変長パケット状態レジスタを更新する。

【0047】

T=0では、入力回線#0-VOQ#0と、入力回線#2-VOQ#1からパケットが読み出される。これら読み出しパケットはいずれも、可変長パケットエンドパケットでないため、可変長パケット状態レジスタを“1”に設定する。(図14の右側参照)これにより、次のスケジューリング周期では、入力回線#0-VOQ#0と、入力回線#2-VOQ#1は、可変長パケット送出中VOQとして扱われ、優先的にスケジューリングされるようになる。

【0048】

次に図15を参照してT=1におけるスケジューリング動作を示す。T=1では、開始入力回線が更新され、入力回線#1→#2→#0の順でスケジューリングが実行される。

● T=1におけるスケジューリング処理

入力回線#1では、全ての可変長パケット状態レジスタ48が“0”であるため、可変長パケット未送出VOQの中から選択を行いVOQ#0を読み出しVOQとする。

【0049】

入力回線#2では、可変長パケット送出中のVOQがあるため、そのVOQを優先的に選択する。この例では、可変長パケット送出VOQはVOQ#1しかないため、VOQ#1を読み出しVOQとする。複数のVOQが存在する場合は、ラウンドロビン方式で選択する。

次に入力回線#0でも、可変長パケット送出中のVOQがあるため、同様に優先的に可変長パケット送出中のVOQの中から読み出しVOQを決定しようとするが、既にVOQ#0は入力回線#1によって選択されているため、選択することが出来ない。従って、可変長パケット未送出回線の中から読み出しVOQを決定する。この例では、VOQ#1とVOQ#2が可変長パケット未送出であり、ラウンドロビンによりVOQ#1を選択する。このように、可変長パケット送出中のVOQがある場合でも、他の入力回線によって選択されている場合で、可変長パケット送出中のVOQが選択できない場合には、可変長パケット未送出VO

Qを選択する。

● T = 1 における可変長パケット状態レジスタ更新

スケジューリング処理によって決定されたVOQからパケットを読み出し、読み出したパケット種別に応じて可変長パケット状態レジスタを更新する。

【0050】

T = 1 では、入力回線 # 0 - VOQ # 1、入力回線 # 1 - VOQ # 0 および入力回線 # 2 - VOQ # 1 からパケットが読み出される。

入力回線 # 0 - VOQ # 1、入力回線 # 1 - VOQ # 0 から読み出されたパケットは共に可変長パケットエンドパケットで無いため可変長パケット送出レジスタを“1”に設定する。これにより、これらのVOQは次のスケジューリング周期では、可変長パケット送出中VOQとして扱われ、優先的にスケジューリングされる。

【0051】

一方、入力回線 # 2 - VOQ # 1 から読み出されたパケットは、可変長パケットエンドパケットであるため、可変長パケット状態レジスタを“0”にクリアする。従って、本VOQは次のスケジューリング周期では、可変長パケット未送出VOQとして扱われる。

このように、可能な限り可変長パケット送出中のVOQを優先的にスケジューリングすることにより、1つの入力回線から送出されるパケットは、VOQ毎に連続して読み出されやすくなるため、出力側の可変長パケット待ち合わせ時間を短くすることが可能となる。

【0052】

図16にこのスケジューリング処理のフローを示す。本フローは、1つの入力回線におけるスケジューリング処理フローである。

ステップ1001～1004はスケジューリングプロセスである。まず、ステップ1001において、優先的に可変長パケット送出中のVOQについてスケジューリングを行う。ここで、未確定とは、他の入力回線によって選択されていないVOQを意味し、要求ありとは、バッファにパケットが蓄積されていることを意味し、可変長パケット送出中とは、可変長パケット状態レジスタに“1”が設

定されていることを意味する。次にステップ1002において、上記ステップ1001の処理によって、読み出しVOQが確定したかを判定し、確定した場合には、可変長パケット状態レジスタの更新プロセスへ、未確定の場合には、可変長パケット未送出VOQのスケジューリングプロセスに移る。

【0053】

ステップ1003は、ステップ1001において、VOQが確定されなかった場合にのみ実行され、可変長パケット未送出（可変長パケット状態レジスタ＝0）のVOQについてスケジューリングを行う。ステップ1004は、上記ステップ1003において、読み出しVOQが確定したかを判定するプロセスであり、未確定の場合にはスケジューリング処理を終了し、確定時は可変長パケット状態レジスタ更新プロセスに移る。

【0054】

ステップ1005～1007は、可変長パケット状態レジスタの更新プロセスである。ステップ1001／1003のスケジューリング処理によって決定されたVOQからパケットを読み出し、読み出したパケット種別を判定し、可変長パケットエンドパケットで無いときには、可変長パケット状態レジスタを“1”に設定し、可変長パケットエンドパケットのときに同レジスタをクリアする。

【0055】

図8、10、11の可変長ベースQoS制御部30ではパケット長を考慮したQoS制御が必要になる。図17は、このパケット長を考慮したQoS制御を実現する手法の一例を概念的に示す。図8の可変長ベースQoS制御部30において、各QoSクラス毎にバッファリングされた先頭のパケットのパケット長に対応するカウンタ値（ここではトークンと呼ぶことにする）を設定する。この際トークン値は、例えば、各QoSクラスに対するウェイト値（ Φ_i ）として式（5）のような正規化した値を使用すれば、パケット長をそのまま使用できる。

【0056】

$$\Phi_i = R \times \tau \times \Phi_{is} / \sum \Phi_{is} \quad \dots (5)$$

ここでRは全体の目標リンクレート、 τ は内部処理の単位時間（ex. 固定長パケット時間）、 Φ_{is} は各QoSクラスに割り当てられたウェイト値、 $\sum \Phi_{is}$ は

アクティブなQoSクラスに対する Φ_{is} 総和である。また添え字の i は、QoSクラス数に対応している。なお、式(5)中の $R \times \tau / \sum \Phi_{is}$ の値は一定であるので、(5)式から計算される Φ_i の代わりに各QoSクラスのウェイト値 Φ_{is} を使うことができる。

【0057】

図17に示すように、各QoSクラスの先頭パケットに対応するトークン値(T_i)を単位時間(τ)毎、 $\Phi_i / \sum \Phi_i$ ずつ減算していき、回線にパケットが読出し可能なとき、各トークン値の中で最小値(ベストエフォートクラス)または0(帯域保証クラス)のトークン値を持つQoSクラスのパケットを読み出す。

【0058】

この手法では、WFQのようにウェイト値に応じて各QoSクラスの帯域をアクティブなQoSクラスで比例分配することが可能である。またWFQのようにキューイングされている全パケットについて、終了予定時刻 $F(i, k)$ を保持しなくて良いため、ハード規模が削減できる。さらに、実時間でスケジューリングするための、読出しレートを規定する帯域保証型のQoSスケジューリングとの間の優先度制御が可能である。また、 $\sum \Phi_i$ をアクティブなキューの数によらず一定とすることにより、そのまま、読出しレートを規定する帯域保証型のQoSスケジューラとして使用できるため、回路を共通化してハード規模の削減や設計工数の削減が図れる。また、トークンを加算する処理とトークンから減算値を減算する処理と各QoSクラスのトークンを比較し、最も小さいまたは0のトークンを持つQoSクラスのキューを選択して読み出す処理は、パイプライン化することも可能であり、高速なQoS制御を行うことも可能である。

【0059】

図18に実際のハードウェア構成の例、図19にフローチャートを示す。図示した例は、可変長ベースQoS制御部において、可変長パケット組立てバッファ部とQoS制御部でパケットバッファは共通でバッファへの書きこみアドレスのみを転送する場合の構成に対応している。出力回線部に入力された固定長パケットは、パケットバッファ50へ書きこまれるとともに、アドレス管理FIFO5

2において、入力回線毎、縮退QoS毎にアドレスのFIFOチェーンを構成する。可変長パケット組立て部54では、各キューへの1可変長パケット分の到着を管理し、1可変長パケット分到着した可変長パケットについて、対応するパケットバッファのアドレスをコピーし、QoSクラス毎、DMUX毎のアドレスFIFOを構成する。さらに、ヘッダ情報をパケット識別部56に転送する。パケット識別部56ではパケットヘッダからQoSクラス、DMUX番号やパケット長といったQoSスケジューリングに必要な情報を抽出し、後で詳述する図19の処理フローに従い、読み出しQoSクラスの選択を行い、DMUX読み出し部58より対応するパケットをパケットバッファから出力パケットハイウェイへ読み出す。このときDMUX読み出し部58では、各DMUX毎の固定的なタイムスロットを管理しておき、各タイムスロットのタイミングにおいてスケジューリングすべきDMUXの番号を通知し、対応するDMUXについて図19の処理フローを実行する。

【0060】

図19の処理フローでは、3つのプロセスが連携して動作する。パケット到着処理プロセスでは、可変長パケットが到着したとき、または前述のアドレス管理FIFO52において、可変長パケットがFIFOの先頭に来たとき、対応する可変長パケットのQoSクラスごとにトークン値の加算、および、ウェイト値総和 $\Sigma \Phi_i$ への加算を行う。またパケット読み出しプロセスでは、前可変長パケットの読み出しが終了し、次の可変長パケットを選択する前、もしくは初期状態において、可変長パケットの選択が行われていないとき、各QoS毎のトークン値(i)を比較し、最小トークン値を持つQoSクラスを選択し、当該QoSクラスのパケットの読み出しを開始する。このとき、同一QoSクラスにキューイングされているパケットが無く、キューが空になるときは、 $\Sigma \Phi_i$ の減算処理を行う。一方、減算値処理プロセスでは、単位時間毎に、全QoSクラスの減算値を計算し、全QoSクラスのトークン値から減算値の減算を行う。

【0061】

図20は、パケット長を考慮したQoS制御の第2の例を概念的に示す図である。トークン値(T_i)をトークン値 $\times \Sigma \Phi_i$ に置き換え、さらに単位時間毎の

減算値として Φ_i を使用する。

この手法では、一般に処理時間がかかる $\Phi_i / \Sigma \Phi_i$ の除算処理をなくすることができる。但し、QoSクラスキューの状態を反映する $\Sigma \Phi_i$ の更新が、前述の手法では1単位時間毎に可能であるのに対して、パケットの先頭を讀出してトークン値を積みあげるときのみになるため、多少公平性に関して特性的な劣化が生じる。

【0062】

図21にフローチャートを示す。パケット到着処理プロセスにおけるトークン値の加算値にパケット長の代わりにパケット長 $\times \Sigma \Phi_i$ を用い、また減算値処理プロセスにおいて、減算値を $\Phi_i / \Sigma \Phi_i$ の代わりに Φ_i としている。

図22は、パケット長を考慮したQoS制御の第3の例を示す。この手法では、トークン値をトークン値 $\times \Sigma \Phi_i$ に置き換える代わりに、 $\Sigma \Phi_i$ の加算をパケット長に相当する回数だけ繰返すことに置き換えている。

【0063】

トークン回数を管理するカウンタが必要となるため多少複雑になるが、乗算処理をなくすることができるためハード量の削減が図れる。またパケット長に相当する回数分の $\Sigma \Phi_i$ の加算の際に、逐次最新の $\Sigma \Phi_i$ を使用することにより、QoSクラスキューの状態を反映する頻度が多くなり、公平性に関して特性的な向上が図れる。

【0064】

図23にフローチャートを示す。パケット到着処理プロセスにおけるトークン値の加算値にパケット長の代わりに $\Sigma \Phi_i$ を用い、累積パケット長カウンタによって、加算した $\Sigma \Phi_i$ の回数を管理する。また減算値処理プロセスにおいて、累積パケット長カウンタが0になるまで、トークン値のフレームへの加算処理を繰返し、減算値を $\Phi_i / \Sigma \Phi_i$ の代わりに Φ_i としている。

【0065】

図24は、最低帯域が保証されさらに余剰帯域の使用が許される最低帯域保証クラスの可変長パケットについてのパケット長を考慮した制御が可能なQoS制御の一例を概念的に示す。入力された最低帯域保証クラスの可変長パケットにつ

いては通常はスイッチ部 1 4 の前段に設けられるポリシング制御部 2 0 において、最低保証帯域を超えて到着したパケットにタギングを行うことにより、帯域保証のパケットとベストエフォート分のパケットが識別される。スイッチ部 1 4 の後段の Q o S 制御部 5 8 では到着した可変長パケットを構成する固定長パケットの数を Q o S クラス (e x 1 6 クラス) 別にカウンタする。Q o S クラスのそれぞれについて帯域保証クラス用のカウンタとベストエフォートクラス用のカウンタが用意されており、最低帯域保証クラスについてはそれらの双方が使用され、ベストエフォートクラスについてはベストエフォートクラスのカウンタのみが使用され、帯域保証クラスについては帯域保証クラスのカウンタのみが使用される。

【 0 0 6 6 】

ここで Q o S 制御部 5 8 では、カウンタ値が正の値のものについて、前述した重み付けによる読出し制御を行う。トークン演算のアルゴリズムは前述の 3 通りのいずれでも良い。但し、この際、可変長パケットのパケット長の代わりに固定長パケットのパケット長を用い、また帯域保証クラスについては設定レートに従う読出しを行う。また帯域保証クラスとベストエフォートクラスからの要求の競合時は単純に帯域保証クラスを高くするかラウンドロビンや予め設定された優先度に応じて読み出す。このようなスケジューリング処理は固定長パケット毎に行い、選択された Q o S クラスのカウンタ値を 1 減算するとともに、可変長バッファ部 6 0 へ選択された Q o S クラスを通知する。

【 0 0 6 7 】

可変長バッファ部 6 0 では、通知数を Q o S クラス毎に計数し、累積数が各キューの先頭のパケット長に達したら、1 可変長パケットの読み出しを行なう。

上記の制御により、1 Q o S クラス内に帯域保証分とベストエフォート分が含まれる最低帯域保証クラスに対して可変長パケットに対する Q o S 制御が可能となる。図 6, 7 を参照して説明した制御のように同一キューにマッピングしてキューしきい値で制御する方法だとベストエフォート分のパケットが利用できるキュー領域は帯域保証クラスが利用できるキュー領域より小さくなるという制限が生じるが、このような問題が解消される。また、単純に 1 Q o S クラスのキュー

を2つに分けると生じるパケットの順序逆転も起きない。

【0068】

さらに、帯域保証クラスのスケジューリングとベストエフォートクラスのスケジューリングを別々に行うことができるので単純に帯域保証クラスの優先度を高くするような優先制御だけでなく、ラウンドロビンにしたり予め優先度を設定しておくなどより複雑な優先制御が可能となる。

図25に具体的な構成を示す。ここでは、固定長パケット単位にパケットバッファへのキューイングを管理しているが、連続する固定長パケットを可変長パケットと置き換え、可変長パケットの先頭と末尾を管理し、1可変長パケット分連続的に読み出すことにより、可変長パケットにおいても全く同様に当てはまる。

【0069】

QoSスケジューラ62において、到着した可変長パケットを構成する固定長パケットの数がカウントされる。この際、当該フレームがタギングされているかいないかによって、適合パケット数カウンタにカウントするか、不適合パケット数カウンタにカウントするかに分かれる。QoSスケジューラ62では各カウンタ値が0で無いものについて、予め設定されたレートやウェイト、優先度に応じて、1つのQoSクラスを選択する。選択されたQoSクラスは可変長パケット管理部64に通知される。可変長パケット管理部64では、通知された選択QoSクラス数を管理しており、累積数をパケット長に換算して、各QoSクラス毎のキューの先頭パケットのパケット長より大きくなった場合には、当該可変長パケットを出力ハイウェイから出力する。

【0070】

このとき、固定長パケットサイズやフォーマットは、どのようなものでもよく、可変長パケットの先頭と末尾を管理できれば、ここでの固定長パケットにヘッダを付与する必要はない。

図26は、最低保証帯域クラスの可変長パケットの制御が可能なQoS制御の他の例の原理を説明する図である。図26はあるQoSクラスに対するトークン値の動作を示している。ここで、入力された可変長パケットはQoSクラスごとにキューイングされている。 Φ_{ig} は帯域保証クラス用に設定された読出しレート

値である。また Φ_{ib} はベストエフォートクラス用に設定されたウェイト値である。ベストエフォート分を含まない帯域保証クラスの Φ_{ib} はゼロに、帯域保証分を含まないベストエフォートクラスの Φ_{ig} はゼロに設定される。最低帯域保証クラスについては、 Φ_{ig} と Φ_{ib} にゼロでない値が設定される。

【0071】

ここでは、1 単位時間内にトークン減算処理を帯域保証クラスサイクル用とベストエフォートクラスサイクル用の 2 回行う。このとき、帯域保証クラスサイクル用のスケジューリングを行い読み出すべきパケットが無いときは、ベストエフォートクラスサイクル用のトークン減算処理も行う。帯域保証クラスサイクル用のスケジューリングを行い読み出すべきフレームがあるときまたは帯域保証クラスの変長パケットの読み出しが続いているときは、ベストエフォートクラスサイクル用のトークン減算処理を行わない。最低帯域保証クラスの packets には予め前段のポリシング制御部において、最低保証帯域を超えたフレームにタギングをしておく。帯域保証クラスサイクル用のスケジューリングを行い読み出すべきフレームの有無を判定する際、タギングのない最低保証帯域以下の packets についてのみ読み出すべき packets が有りとして処理し、タギングがある場合は読み出すべき packets が無しとして処理する。

【0072】

もしくは、帯域保証クラスサイクル用のスケジューリングを行ったときの読み出すべき packets の有り無しに関わらず、常にベストエフォートクラスサイクル用のトークン減算処理を行い、 $\Sigma \Phi_{ib}$ の代わりに $\Sigma \Phi_{iall}$ (アクティブなコネクションについての Φ_{ig} と Φ_{ib} の総和) を使用することもできる。

このようにすれば、変長 packets の状態のままで、スケジューリングができ、帯域保証クラスのスケジューリングを行ってからベストエフォートクラスのスケジューリングを行うという簡単な制御で最低帯域保証クラスの収容が可能となる。但し、一般に $\Sigma \Phi_{ig}$ と $\Sigma \Phi_{ib}$ は異なる値であるため、前記の 3 通りのトークン演算アルゴリズムのうち、トークンとしてどちらのクラスにも共通であるフレーム長を用いる第 1 番目の演算アルゴリズムのみが適用できる。

【0073】

図 2 7 に処理のフローチャートを示す。フレーム到着処理プロセスにおいて、帯域保証クラスとベストエフォートクラスで、 $\Sigma \Phi_i$ を各々 $\Sigma \Phi_{ig}$ と $\Sigma \Phi_{ib}$ として別々に管理する。また同様に減算値更新プロセスにおいても減算値 $\Phi_i / \Sigma \Phi_i$ の代わりに $\Phi_{ig} / \Sigma \Phi_{ig}$ と $\Phi_{ib} / \Sigma \Phi_{ib}$ を別々に管理し、トークンの減算処理を帯域保証クラスからフレームが読み出されているときと、そうでないときに分けて行う。またフレーム読み出しプロセスにおいて、先に帯域保証クラスのキューイングされているフレームの選択を行い、次にベストエフォートクラスのキューイングされているフレームの選択を行う。また $\Sigma \Phi$ の減算についても同様に帯域保証クラスとベストエフォートクラスで別々に行う。なお、図 2 7 に示した例では、すべての QoS クラスについて帯域保証クラス用とベストエフォートクラス用の 2 つのトークンのカウンタが用意され、帯域保証クラスについては、 Φ_{ib} がゼロに設定され、ベストエフォートクラスについては、 Φ_{ig} がゼロに設定されるものとする。

【0074】

図 2 8 は、最低帯域保証クラスの可変長パケットの制御が可能な QoS 制御のさらに他の例の原理を説明する図である。図 2 8 は図 2 6 と同様に前述の第 1 番目の演算アルゴリズムを適用した場合の、ある QoS クラスに対するトークン値の動作を示している。この例では、最低帯域保証クラスのパケットは、帯域保証クラス用とベストエフォートクラス用の 2 つのスケジューラにより並列に処理する。但し、トークン値は QoS クラス毎のキューの先頭パケットのパケット長に対応する同一のトークン値を使用する。このとき、ベストエフォートクラス用のスケジューラでは、 $\Sigma \Phi_{ib}$ の代わりに $\Sigma \Phi_{i_{all}}$ を使用する。

【0075】

最低帯域保証クラスについては、帯域保証クラス用とベストエフォートクラス用の 2 つのスケジューラによるスケジューリングが行なわれる。それらの読出し処理により、いずれか先に読出しが行われたときは、読出しが行われなかった方の読み出されたフレームに対して行われた累積トークン減算値を繰り越し分として、次のフレームのトークン値を計算する際に減算する。

【0076】

この手法では、帯域保証クラス用とベストエフォートクラス用で別々にトークンの減算処理が必要となるため、ハード規模が増加する。但し、帯域保証クラス用とベストエフォートクラス用を並列に処理できるので、高速なスケジューリングが可能となる。また、 $\Sigma \Phi_{ig}$ と $\Sigma \Phi_{iall}$ が一致する必要がないので、前述した3つの演算アルゴリズムのいずれもが適用可能である。

【0077】

図29にフローチャートを示す。この例では、フレーム到着処理プロセスにおいて、トークン値を帯域保証クラスとベストエフォートクラスでトークン値(i_g)、トークン値(i_b)として別々に管理する。さらに、各々のトークン値への加算時に可変長パケット読出プロセスにおいて可変長パケットを選択する際、選択されなかったクラスの当該パケットに対する累積の減算値を繰越値として保持しておき、その値をパケット長から減算した値をトークン値とする。また $\Sigma \Phi_i$ の代わりに帯域保証クラスの Φ_i のみを対象として $\Sigma \Phi_{ig}$ とすべてのQoSクラスを対象として $\Sigma \Phi_{iall}$ を別々に管理する。また減算値として $\Phi_i / \Sigma \Phi_i$ の代わりに $\Phi_{ig} / \Sigma \Phi_{ig}$ と $\Phi_{ib} / \Sigma \Phi_{iall}$ を別々に管理し、対応するトークン値に対して減算を行う。可変長パケット読み出しプロセスにおいて、先に帯域保証クラスのキューイングされている可変長パケットの選択を行い、次にベストエフォートクラスのキューイングされている可変長パケットの選択を行う。このとき選択されなかったクラスについては前述の繰越値を保持しておき、トークン値更新時にトークン値から減算する。図29の例でも図27の例と同様、すべてのQoSクラスについて、2つずつのトークン値のカウンタが用意される。

【0078】

次に、本発明の一実施形態に係るパケット廃棄の手法を説明する。なお、以下の記述において、「可変長パケット」は「フレーム」と、「固定長パケット」は単に「パケット」と称される。本発明において、廃棄はランダムでなく周期的に行い、その周期をキュー長の状態に応じて変えることにより、確率的な廃棄を実現する。また、廃棄を読み出し時に行うことにより、キューの先頭から廃棄することが出来るため、廃棄を早く端末に通知することが出来る。廃棄判定はキューの先頭パケット読み出し時に行う。図30および式(6)に示すように、キュー

長に応じた廃棄確率を決定し、式(7)で決定される周期で廃棄を行う。これにより、長いフレームはそれだけ高い確率で廃棄することが出来る。廃棄時は、1フレーム分のパケットを1パケット時間に廃棄することにより、スループットの低下を防ぐことが出来る。

【0079】

$$\text{廃棄確率} = (\text{平均キュー長} - TH_{\min}) / (TH_v - TH_{\min}) \quad \cdots (6)$$

$$\text{廃棄周期} = 1 / (\text{廃棄確率} \times \text{パケット長}) \quad \cdots (7)$$

具体的な回路構成を図31に、動作フローを図32に、動作の一例を図33に示す。バッファ内のキュー長を管理するキュー長カウンタ66、設定されたしきい値を保持するしきい値レジスタ68、設定された読み出し間隔パラメータを保持する読み出し間隔レジスタ70、読み出し間隔制御に用いるLBカウンタ72、バッファの先頭のフレームのフレーム長を管理するフレーム長カウンタ74、LBカウンタ動作を制御するカウンタ制御76、その他演算器より構成される。

【0080】

フレームの先頭のパケットを読み出す時、平均キュー長としきい値を比較する。平均キュー長がしきい値を越えていない場合はそのまま読み出しを行う。しきい値を越えた場合、しきい値と平均キュー長との差分にフレーム長を掛け合わせた値を、LBカウンタ72に加算する。そして加算結果が0を越えていた場合、該当フレームに属するパケットを全て廃棄し(具体的な方法は後述)、LBカウンタ値から、読み出し間隔レジスタ70の値にフレーム長を掛け合わせた値を減算する。加算結果が0を越えていない場合は、そのまま読み出しを行う。LBカウンタに加算する値が大きいほど、カウンタ値が0を越える確率が高くなる。よって本動作により、輻輳度合いとフレーム長に応じた確率でフレームを廃棄することが出来る。

【0081】

ここで、平均キュー長はその時点での値を用いることも、移動平均を取った値を用いることも可能である。一時的な輻輳ではなく、長期的な輻輳時のみ廃棄制御をかけるため、移動平均を取った値を用いることが有効であるが、バッファを大きく取ることができれば、しきい値を大きく設定し実キュー長がしきい値越え

となった時点で長期輻輳と判断することも可能である。また、読み出し時に廃棄することで、バッファの先頭のフレームに対し廃棄することが出来る。

【0082】

図34および図35に1フレームに属するすべてのパケットの廃棄を1パケット時間内に行なう手法の一例を示す。図の左上に概念的に表されるアドレス管理FIFOは、実際には各キューのスタート/エンドポインタ92とアドレス管理テーブル94により構成される。アドレス管理テーブル94には、次のアドレスを示す次アドレス、フレーム最終パケットのアドレスを示すENDアドレス、キューに次のフレームが存在することを示す次先頭アドレスイネーブル、次のフレームの先頭アドレスを示す次先頭アドレス、フレームの先頭を示すフレーム先頭フラグ、フレーム長のフィールドを有するレコードが格納される。

【0083】

図34において、廃棄するフレームを空きアドレスキューにつなぎ替えるため、空きアドレスキューのエンドポインタによって示される最終アドレス（アドレス9）のレコードの次アドレスの値をつなぎ替えるフレームの先頭アドレス（アドレス5）に書き替える（ステップA）。そして空きアドレスキューのエンドポインタをつなぎ替えるフレームのENDアドレス値（アドレス1）に書き替える（ステップB）。次に廃棄が行われたキューから廃棄フレームを取り除くため、キューのスタートポインタを次先頭アドレス（アドレス13）に書き替える（ステップC）。以上のようにアドレスリンクを書き替えることにより図35に示すようになり、1パケット時間内に1フレーム分の廃棄を行うことが出来る。

【0084】

パケット廃棄方式の他の例として、基本的な考えは前述と同じであるが、廃棄判定は各パケット読み出し毎に行い、廃棄確率は、平均キュー長がしきい値を越えた量のみに比例させ、それに見合った周期で廃棄を行う。パケット毎に廃棄判定を行うため、結果的に長いフレームはそれだけ高い確率で廃棄することが出来る。1パケット廃棄したら同一フレームに属する以降のパケットの読み出し毎に廃棄する。廃棄に廃棄パケット数分の時間掛かるため特性的に前述の手法に劣るが、逆にパケットbyパケットの動作のため回路構成が簡単になるため、より高

速動作が必要なバッファ制御に適する。

【0085】

以上の廃棄方式はキュー毎に制御される。また、複数のしきい値を持たせることにより廃棄に対するプライオリティを持たせることが出来る。

このパケット廃棄方式の具体的な構成の例を、図36に示す。キュー長を管理するキュー長カウンタ78、設定されたしきい値を保持するしきい値レジスタ80、設定された読み出し間隔パラメータを保持する読み出し間隔レジスタ82、読み出し間隔制御に用いるLBカウンタ84、フレーム廃棄中であることを示す廃棄フラグ86、LBカウンタを制御するカウンタ制御部88、その他演算器より構成される。各パケットの読み出し時、平均キュー長としきい値を比較する。平均キュー長がしきい値を越えていない場合はそのまま読み出しを行う。しきい値を越えた場合、しきい値とキュー長との差分を、LBカウンタに加算する。そして加算結果が0を越えていた場合、該当パケットを廃棄、廃棄LBカウンタ値から読み出し間隔レジスタ値を減算し、廃棄フラグを立てる。加算結果が0を越えていない場合は、そのまま読み出しを行う。パケット廃棄を行った場合、同一フレームの以降のパケット読み出し時に廃棄を行い、LBカウンタ値に対し、LBしきい値とキュー長との差分値の加算と、カウンタ値から読み出し間隔レジスタ値を減算する。同一フレームの最終になったら、フラグをクリアする。（先頭フレーム到着時、廃棄フラグをクリアしてから廃棄判定をしても良い。）最終パケットは廃棄しないことで、廃棄フレームと次のフレームとを識別させることが出来る。パケット毎に廃棄判定を行うため、廃棄確率はフレームが長いほど高くなる。よって本動作により、輻輳度合いとフレーム長に応じた確率でフレームを廃棄することが出来る。以上の動作フローを図37に、LBカウンタ動作例を図38に示す。ここで、平均キュー長はその時点での値を用いることも、移動平均を取った値を用いることも可能である。一時的な輻輳ではなく、長期的な輻輳時のみ廃棄制御をかけるため、移動平均を取った値を用いることが有効であるが、バッファが大きく取れば、しきい値を大きく設定し実キュー長がしきい値越えとなった時点で長期輻輳と判断することも可能である。また、読み出し時に廃棄することで、バッファの先頭のフレームに対し廃棄することが出来る。パケット

の廃棄に1フレーム時間かかるため、前述の方式より特性的には劣化するが、回路構成が簡単になるため、より高速な動作が必要なところに適する。

【0086】

LBカウンタ制御については、正負を逆転したり、廃棄判定を0でなくしきい値を用いて行っても良い。

可変長パケットであるフレームをスイッチングするパケットスイッチにおいては、1つのフレームを複数の低速回線へマルチキャストする場合に次の様な解決すべき問題がある。すなわち、図39に示すように、スイッチ部16からフレーム組立バッファ28に1フレーム分のパケットが溜まった後にDMUX90にパケットが移し替えられるが、その移し替えを1パケットずつ行くと、マルチキャストを行う場合、図40に示すように複数の低速回線に同時に複製するためにはマルチキャスト数分のアドレス更新を1パケット時間内に行う必要があり高速動作には適さない。また図41に示すように、マルチキャスト用に別のキューを設ける場合、フレーム連続性を保証するためにマルチキャストを行なわないパケットとマルチキャストの複製先が競合しないように制御する必要がある。更に、フレーム長に比例してマルチキャストに時間がかかる問題もある。

【0087】

そこで、図42に示すように、フレーム組立てバッファ28からDMUX90への移動を1パケットずつでなく1フレーム分のパケットをまとめて移動するようにすれば、フレーム組立てバッファ28での読み出し競合の問題は解消される。1フレーム分のパケットの移動は、以下に説明するように、パケットのデータを実際に移動するのではなく、パケットバッファ内の格納アドレスを移動するようにすれば、1パケット時間内で可能になる。

【0088】

フレーム単位でのパケットの移動の具体例を図43、44を参照して説明する。図の左上に概念的に示されるアドレス管理FIFOは、実際には、各キューのスタート/エンドポインタ92とアドレス管理テーブル94により構成される。アドレス管理テーブル94には、次のアドレスを示す次アドレス、フレーム最終パケットのアドレスを示すENDアドレス、キューに次のフレームが存在するこ

とを示す次先頭アドレスイネーブル、次のフレームの先頭アドレスを示す次先頭アドレス、フレームの先頭を示すフレーム先頭フラグ、フレーム長のフィールドを有するレコードが格納される。

【0089】

ここで、図43において、1フレーム分のパケットがフレーム組立バッファに溜まった場合、まずフレームを出力バッファキューにつなぐために、出力バッファのエンドポインタが示す最終アドレス（アドレス2）のレコードの次アドレスの値を、フレームの先頭アドレス（アドレス5）に書き替える（ステップA）。そして出力バッファのエンドポインタをつなぎ替えるフレームのENDアドレス値（アドレス1）に書き替える（ステップB）。次にフレーム組立キューを更新するために、フレーム組立キューのスタートポインタを次先頭アドレス（アドレス13）に書き替える（ステップC）。以上により、図44に示すようになり、1パケット時間内にフレーム組立バッファから出力バッファに移動することが出来る。

【0090】

図45に本発明におけるマルチキャストの処理を概念的に示す。マルチキャストすべきフレームはまずDMUX90を経てマルチキャストキュー96に移動される。勿論、この移動もアドレスの書き替えのみによって行なわれる。マルチキャストキュー96からDMUX90の各出力回線に複製する際、複製フレーム毎に1つのマルチキャストアドレス98を付与する。そしてアドレスリンクにはマルチキャストアドレス98をつなぎ変える。これにより、1フレーム分を1パケット時間で仮想的に複製できるため、ユニキャストパケットとのフレーム連続性が保証され競合制御が不要になる。同時に複製に掛かる時間が複製数×パケット時間のみで可能となる。

【0091】

マルチキャスト処理の具体例を図46～図53に示す。ここでは、出力バッファは更に複数の出力回線に対するキューを含む。図43、44の構成に加え、複数の出力回線へのキューとマルチキャストキュー96のスタート／エンドアドレスポインタとコネクション管理テーブル100を用意する。また、アドレス管理

テーブル94には、マルチキャスト識別ビット、マルチキャストコネクションを識別するCID、マルチキャストキューからの再配置完了を示す振分完了フラグ、フレームのマルチキャスト出力数を管理するマルチキャスト数カウンタ、マルチキャストアドレスに対応するフレームのアドレスを示すフレームアドレスのフィールドが追加される。コネクション管理テーブルには、マルチキャストされるコネクションのCIDが実際の出力回線の番号とともに格納される。図示されている例は、CIDの2と1がこの順で登録され、それらに対応する出力回線はそれぞれ1と4であることを示す。

【0092】

マルチキャストフレームのフレーム組立が完了すると、出力バッファではなくマルチキャストキューに移動する。出力回線別のキューに同一のアドレスを使用できないため、複製時に新たな空きアドレス（アドレス3，11）を確保し、そのアドレスをフレームの代理として該当する出力バッファキュー（キュー1，4）につなぎ変える。同時にアドレス管理テーブルのマルチキャストアドレス（アドレス3，11）に対し、フレームアドレスに実際のフレームのスタートアドレス（アドレス5）を、そのアドレスのENDアドレスにフレームのエンドアドレス（アドレス1）を保持する。また、フレームがコピーされる毎にそのアドレスのマルチキャストカウンタをインクリメントし、同一フレームの最終コピー時（コネクション管理テーブルのENDフラグで識別）に再配置完了フラグを立てる。

【0093】

DMUXからの読み出し時は、マルチキャストアドレス（アドレス3，11）のフレームアドレス（アドレス5）からENDアドレス（アドレス1）までのパケットを読み出す。1フレーム分のパケット読み出し完了後、マルチキャストアドレスを空きアドレスキューに返却し、マルチキャストカウンタをデクリメントする。その結果が0で且つ再配置完了フラグが立っていれば、リンクとポインタの書き換えにより1フレーム分のアドレスを空きアドレスに返却する。ここで、第一回目のコピーではマルチキャストアドレスを用いずリンクとエンドポインタの書き換えにより再配置することで、制御が複雑になるものの使用アドレスを削

減することも出来る。

【0094】

【発明の効果】

以上説明したように本発明によれば、高速かつ少ないハードウェア規模で、可変長パケットのQoS制御、廃棄制御およびマルチキャスト制御を実現する大容量パケットスイッチが提供される。

【図面の簡単な説明】

【図1】

本発明の一実施形態に係る大容量パケットスイッチを概念的に示す図である。

【図2】

可変長パケットの一例としてのIPv4のフレーム形式を示す図である。

【図3】

固定長パケットの形式の一例を示す図である。

【図4】

スイッチ部16の入力例におけるスケジューリングを実現する具体的な構成を示す図である。

【図5】

パイプライン処理によるスケジューリングの一例を示す図である。

【図6】

入力側における最低帯域保証クラスのパケットのスケジューリングの一例を示す図である。

【図7】

図6のスケジューリング処理のフローチャートである。

【図8】

出力バッファ部18の構成の第1の例を示す図である。

【図9】

出力バッファ部18の構成の第2の例を示す図である。

【図10】

図8の例の詳細な構成を示す図である。

【図 1 1】

図 8 および図 1 0 の構成において、出力回線がさらに複数の低速回線へ分離される場合を示す図である。

【図 1 2】

図 9 の構成において、出力回線がさらに複数の低速回線へ分離される場合を示す図である。

【図 1 3】

出力側の処理を考慮した入力側のスケジューラの構成を示す図である。

【図 1 4】

図 1 3 のスケジューラの動作の一例を示す図である。

【図 1 5】

図 1 3 のスケジューラの動作の一例を示す図である。

【図 1 6】

図 1 3 のスケジューラの動作のフローチャートである。

【図 1 7】

本発明における可変長パケットの Q o S 制御の第 1 の例を概念的に示す図である。

【図 1 8】

図 1 7 の Q o S 制御を実現する回路構成の一例を示す図である。

【図 1 9】

図 1 7 の Q o S 制御の動作のフローチャートである。

【図 2 0】

可変長パケットの Q o S 制御の第 2 の例を概念的に示す図である。

【図 2 1】

図 2 0 の Q o S 制御の動作のフローチャートである。

【図 2 2】

可変長パケットの Q o S 制御の第 3 の例を概念的に示す図である。

【図 2 3】

図 2 2 の Q o S 制御の動作のフローチャートである。

【図 2 4】

最低帯域保証クラスの変長パケットの制御が可能な Q o S 制御の第 1 の例を示す図である。

【図 2 5】

図 2 4 の Q o S 制御を実現する具体的な構成を示す図である。

【図 2 6】

最低帯域保証クラスの変長パケットの制御が可能な Q o S 制御の第 2 の例を示す図である。

【図 2 7】

図 2 6 の Q o S 制御の動作のフローチャートである。

【図 2 8】

最低帯域保証クラスの変長パケットの制御が可能な Q o S 制御の第 3 の例を示す図である。

【図 2 9】

図 2 6 の Q o S 制御の動作のフローチャートである。

【図 3 0】

本発明のパケット廃棄制御における平均キュー長と廃棄確率の関係を示す図である。

【図 3 1】

本発明のパケット廃棄制御の第 1 の例の構成を示すブロック図である。

【図 3 2】

図 3 1 の廃棄制御の処理のフローチャートである。

【図 3 3】

図 3 1 の廃棄制御における L B カウンタの動作の一例を示すグラフである。

【図 3 4】

本発明の廃棄制御の動作の一例を示す図である。

【図 3 5】

本発明の廃棄制御の動作の一例を示す図である。

【図 3 6】

本発明の packets 廃棄制御の第 2 の例の構成を示すブロック図である。

【図 3 7】

図 3.6 の packets 廃棄制御の処理のフローチャートである。

【図 3 8】

図 3.6 の packets 廃棄制御における LB カウンタの動作の一例を示す図である。

【図 3 9】

可変長 packets のマルチキャスト制御において解決すべき課題を説明する図である。

【図 4 0】

可変長 packets のマルチキャスト制御において解決すべき課題を説明する図である。

【図 4 1】

可変長 packets のマルチキャスト制御において解決すべき課題を説明する図である。

【図 4 2】

本発明における、フレーム組立てバッファから DMUX のフレーム単位での移動を概念的に示す図である。

【図 4 3】

フレーム単位での移動の具体例を示す図である。

【図 4 4】

フレーム単位での移動の具体例を示す図である。

【図 4 5】

本発明の一実施例に係るマルチキャスト制御を概念的に示す図である。

【図 4 6】

マルチキャスト制御の具体例を示す図である。

【図 4 7】

マルチキャスト制御の具体例を示す図である。

【図 4 8】

マルチキャスト制御の具体例を示す図である。

【図 4 9】

マルチキャスト制御の具体例を示す図である。

【図 5 0】

マルチキャスト制御の具体例を示す図である。

【図 5 1】

マルチキャスト制御の具体例を示す図である。

【図 5 2】

マルチキャスト制御の具体例を示す図である。

【図 5 3】

マルチキャスト制御の具体例を示す図である。

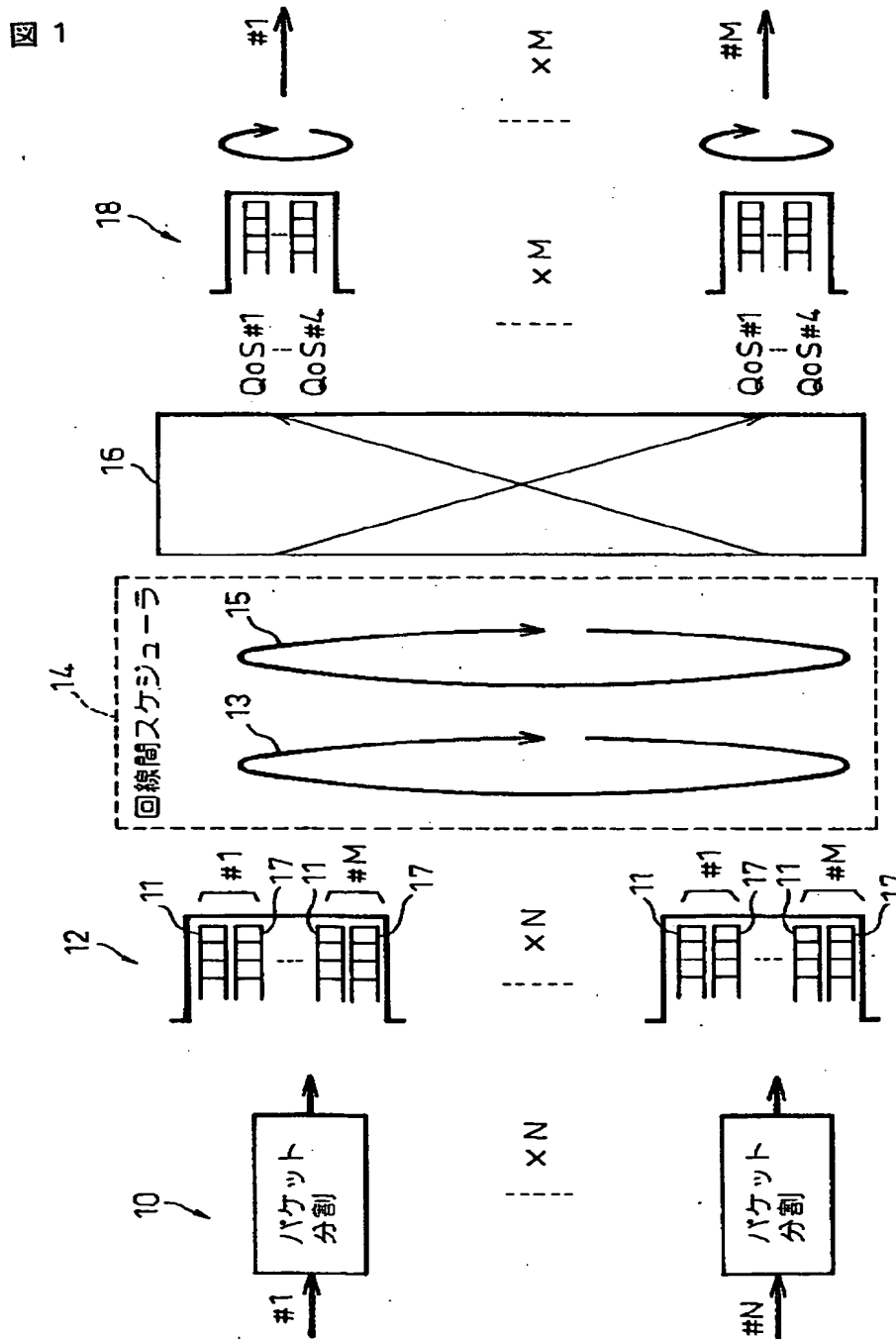
【符号の説明】

- 1 0 … パケット分割部
- 1 2 … 入力バッファ部
- 1 4 … 回線間スケジューラ
- 1 6 … スイッチ部
- 1 8 … 出力バッファ部

【書類名】

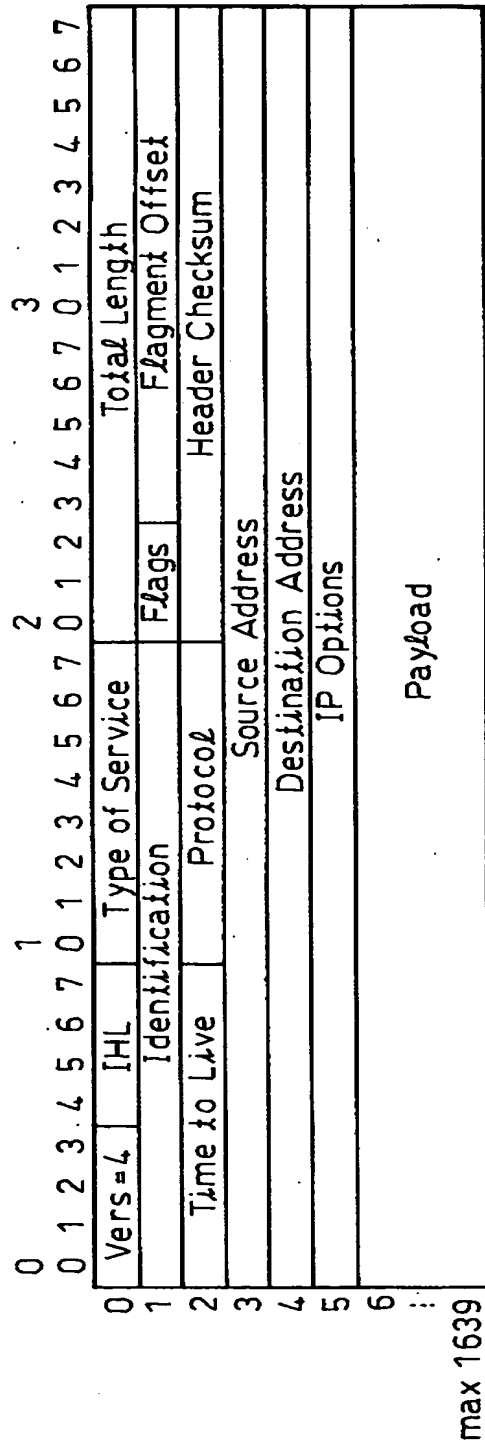
図面

【図 1】



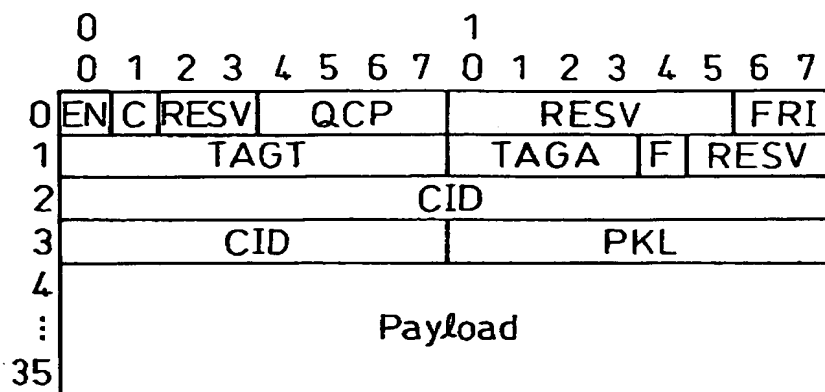
【図 2】

図 2

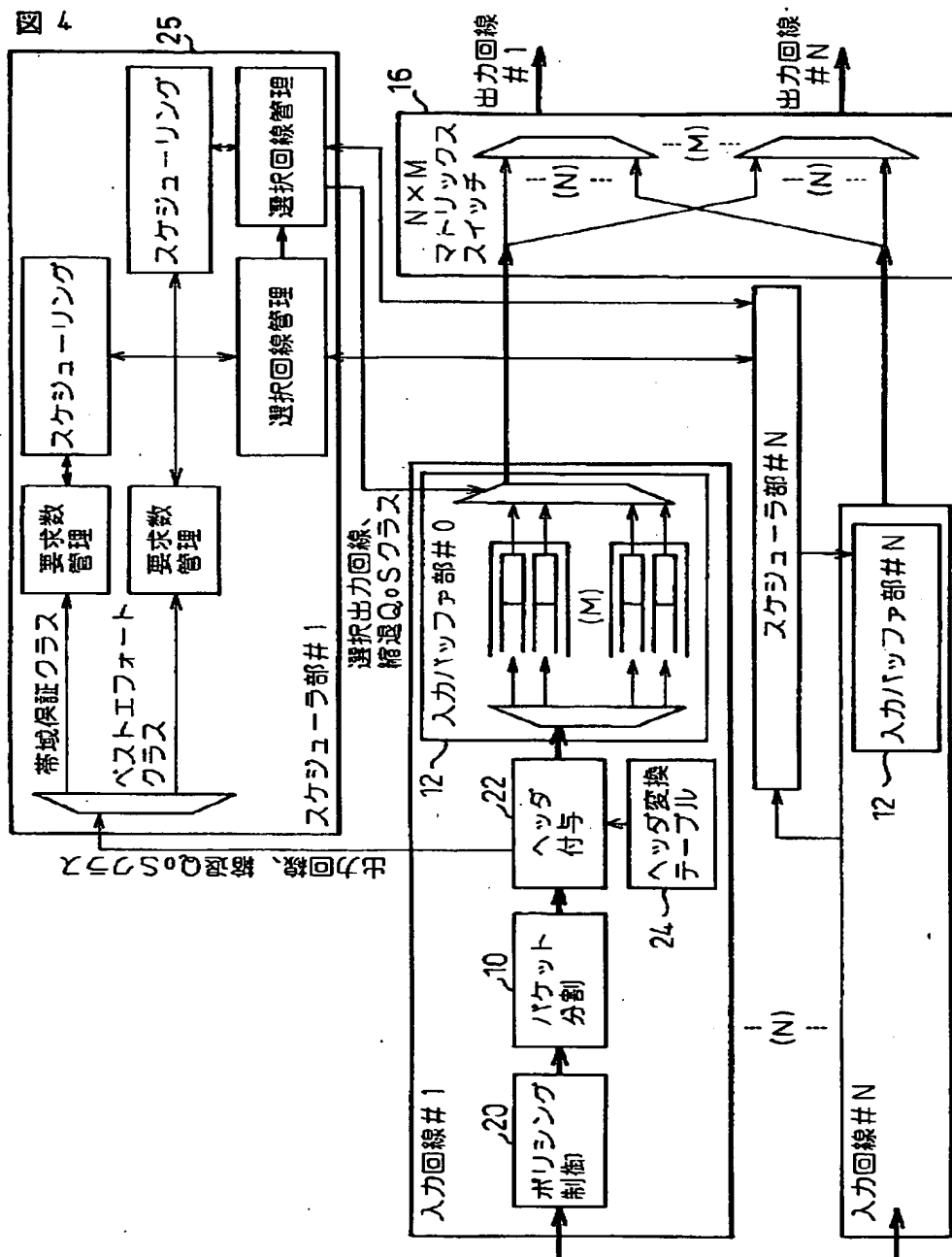


【図 3】

図 3

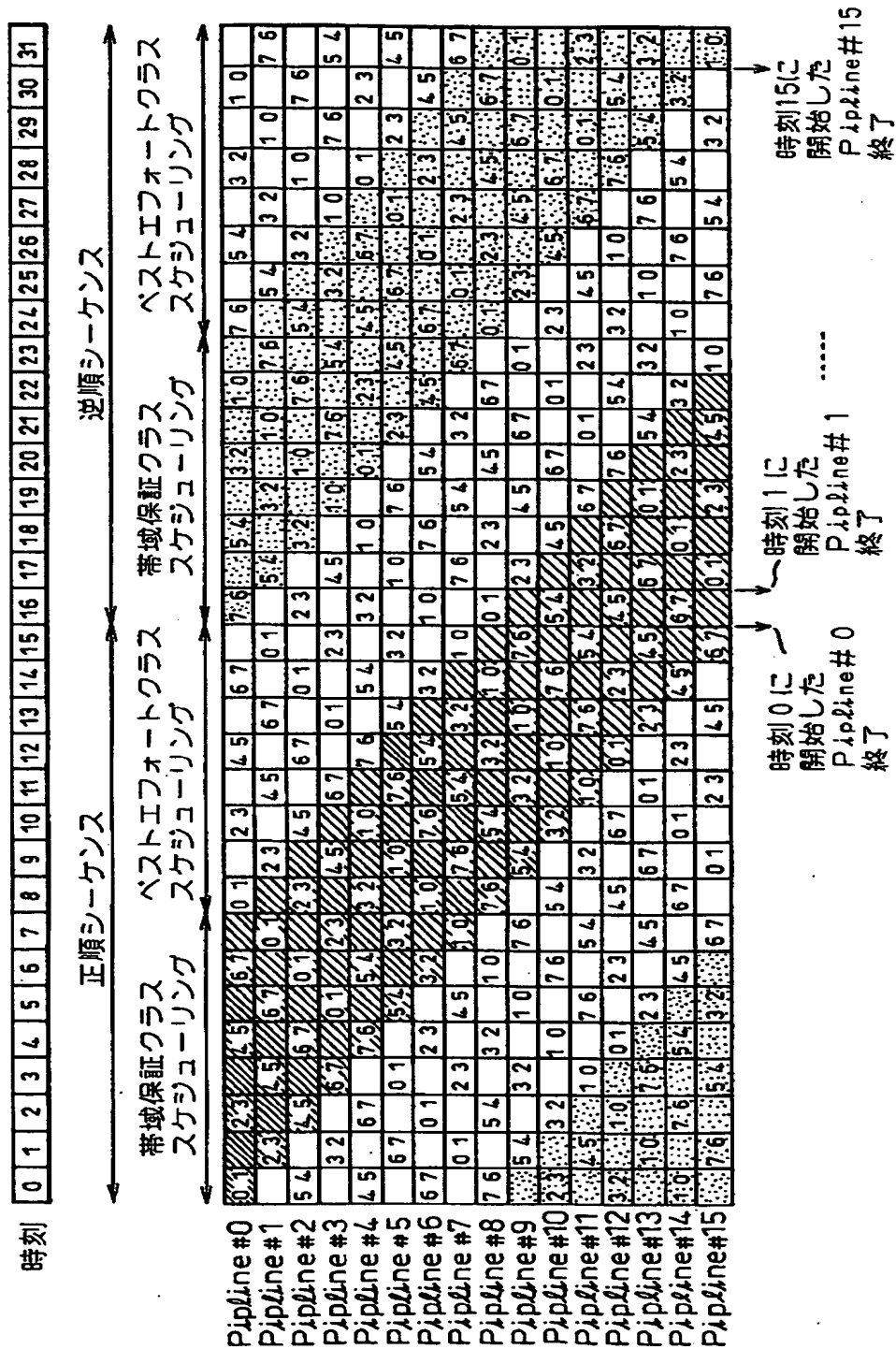


【図4】



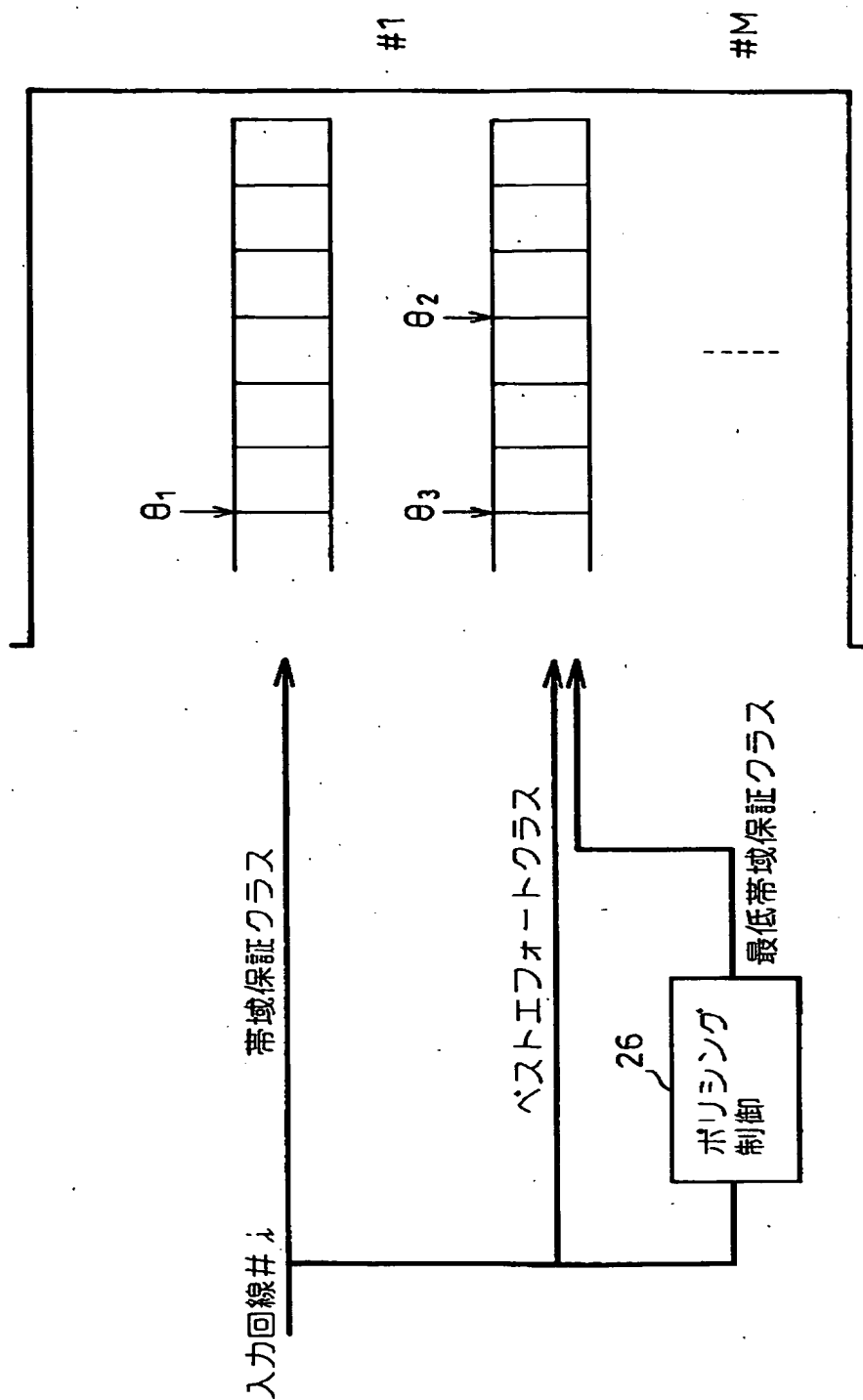
【図 5】

図 5



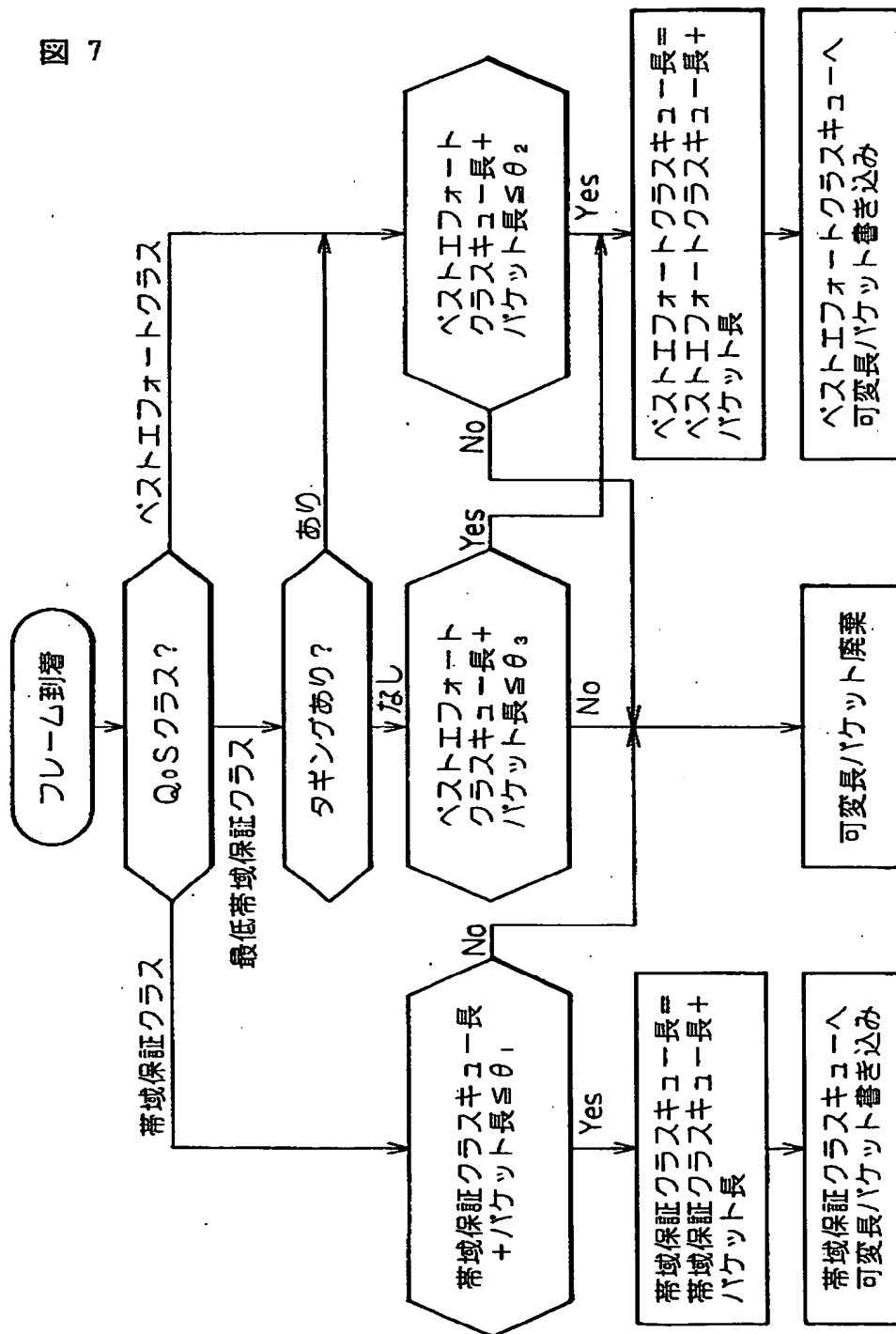
【図 6】

図 6



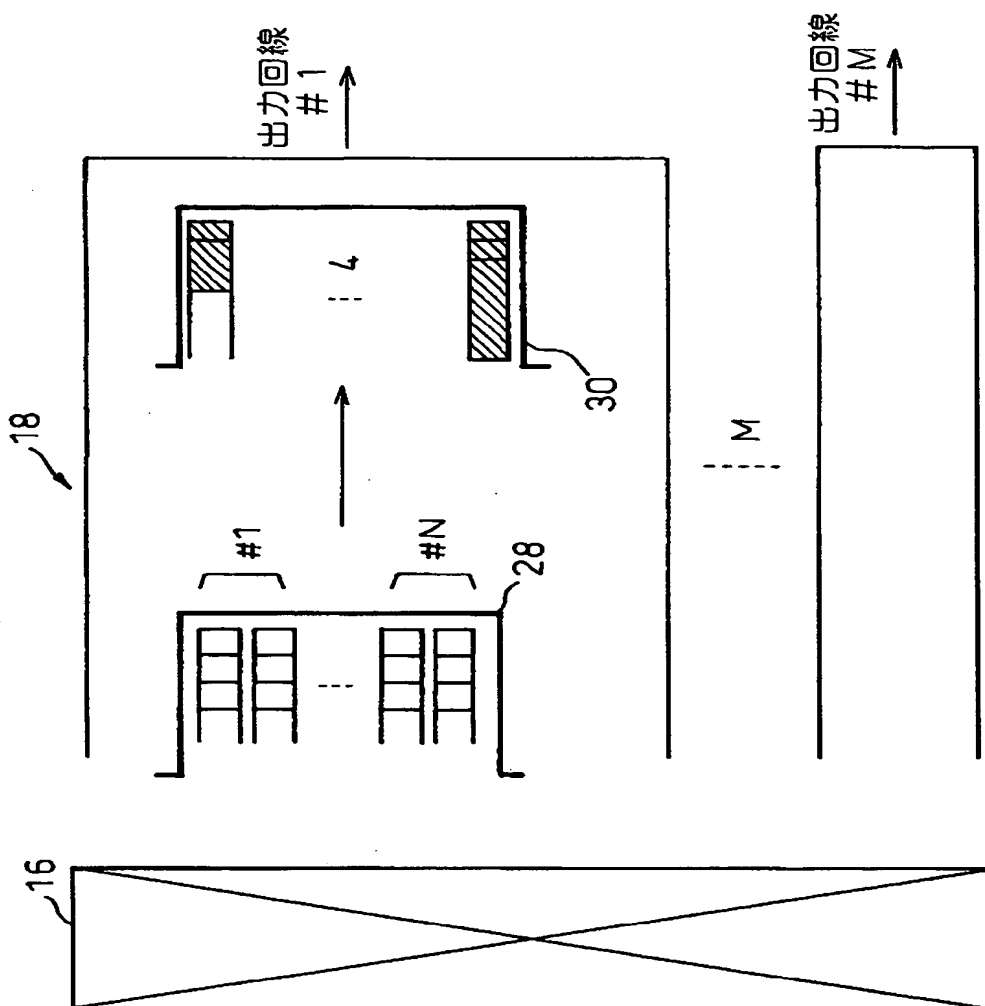
【図 7】

図 7



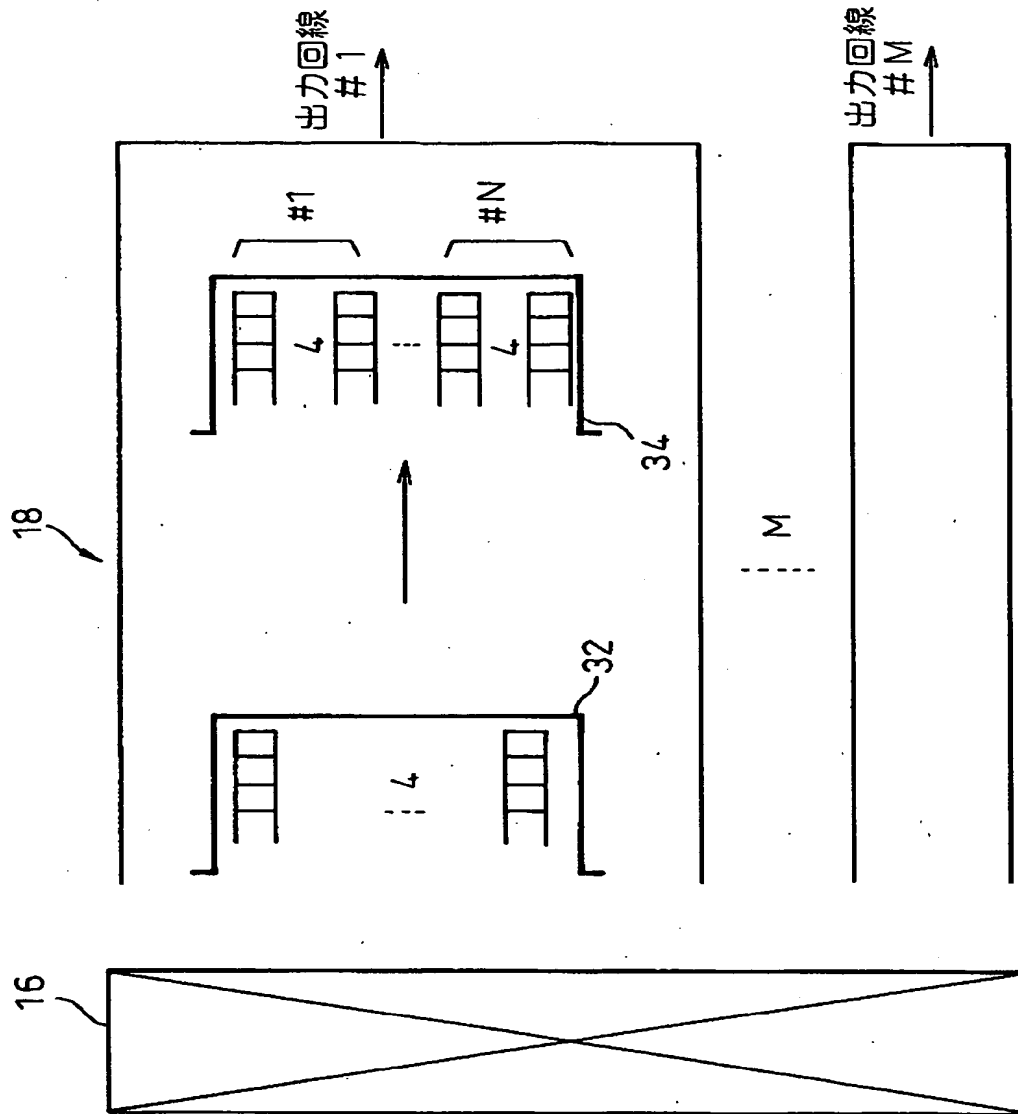
【図 8】

図 8



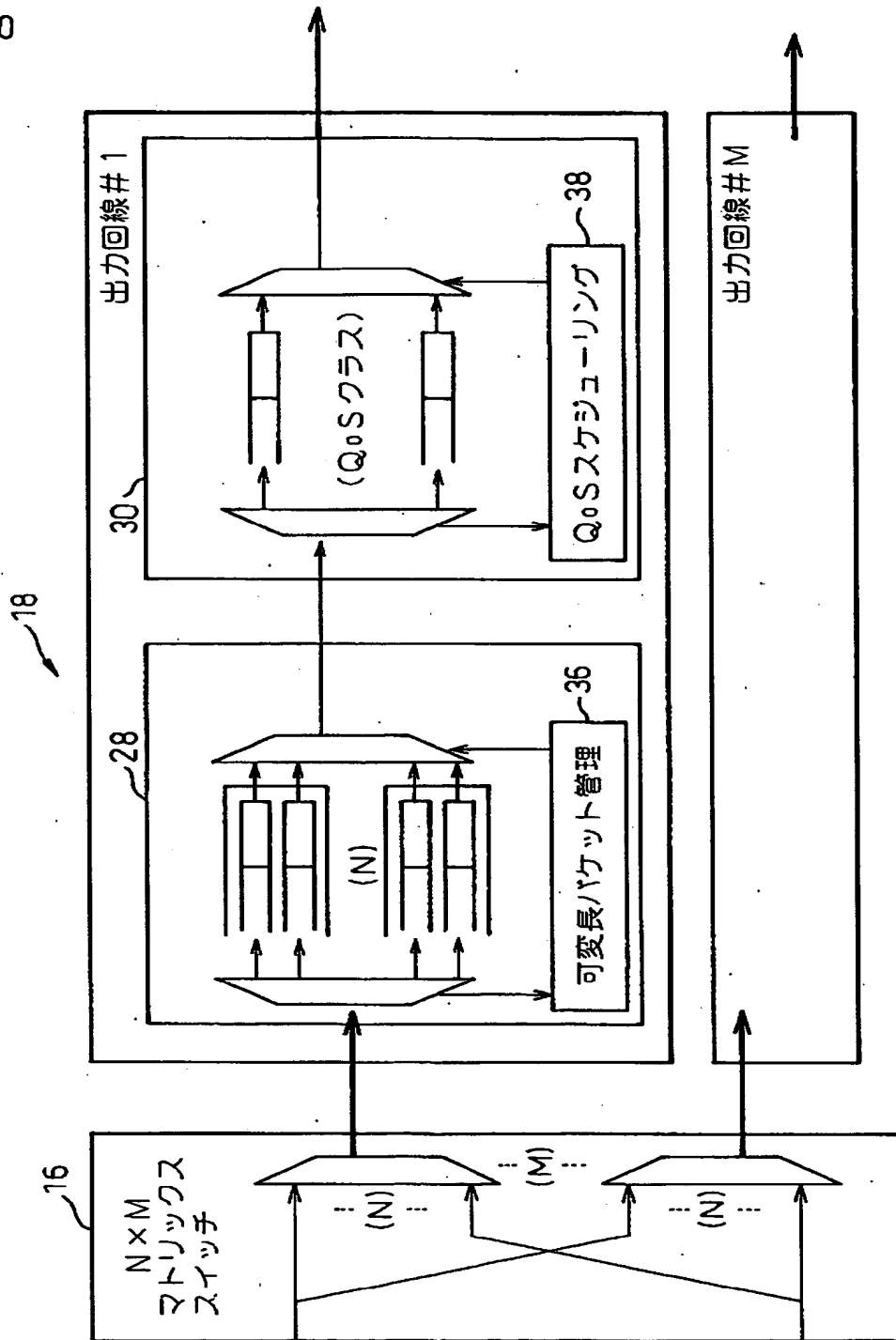
【図9】

図 9

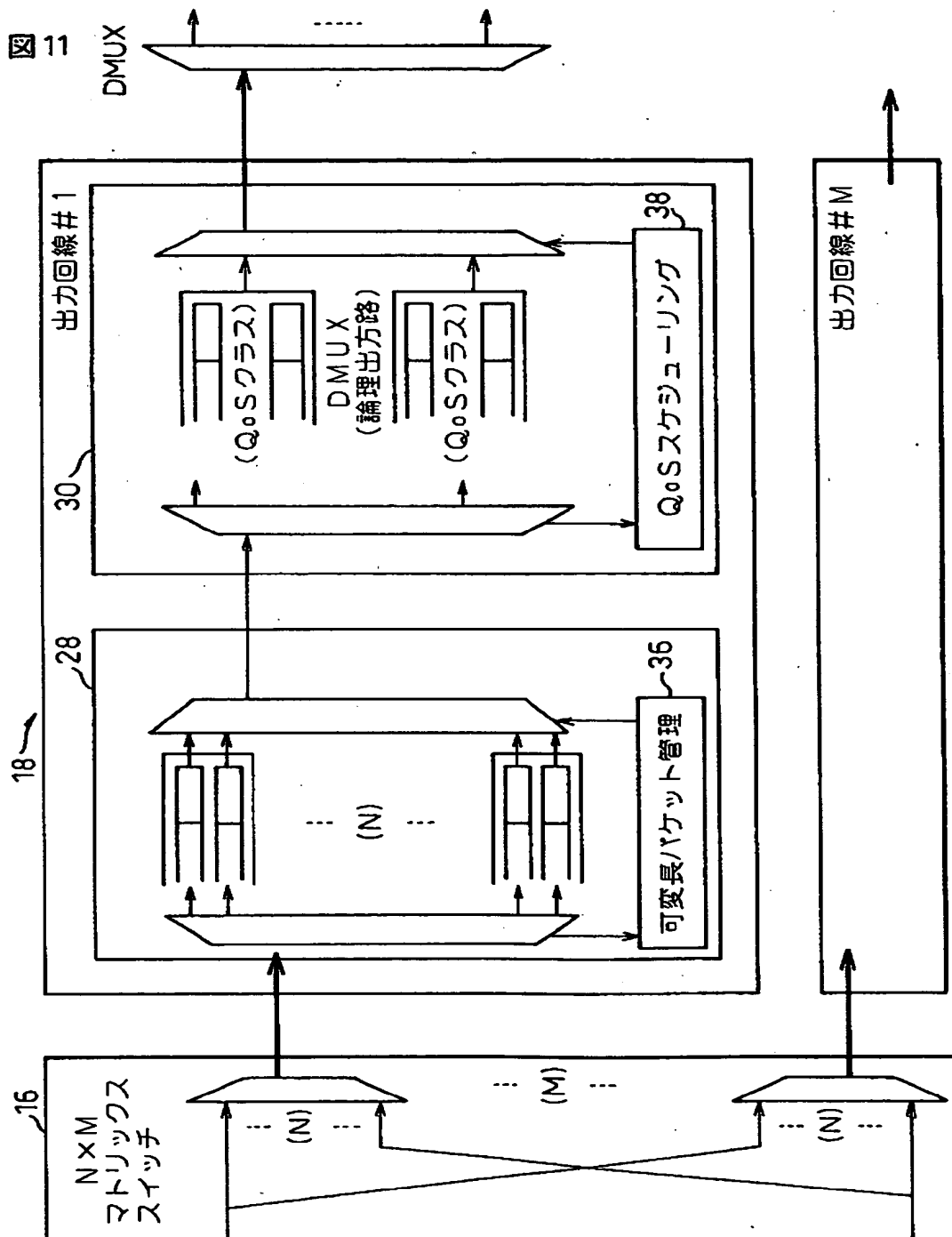


【図10】

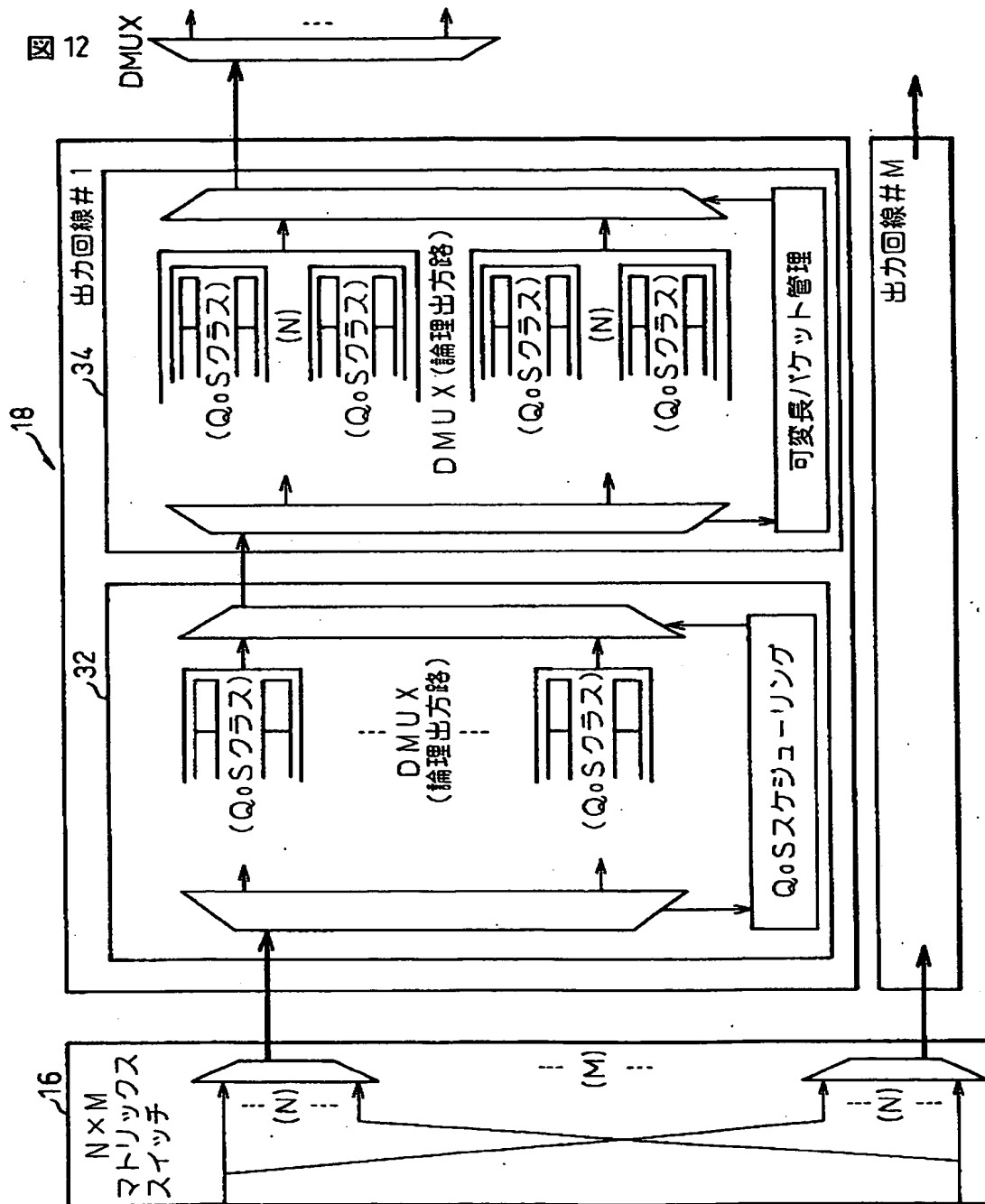
図 10



【図 11】

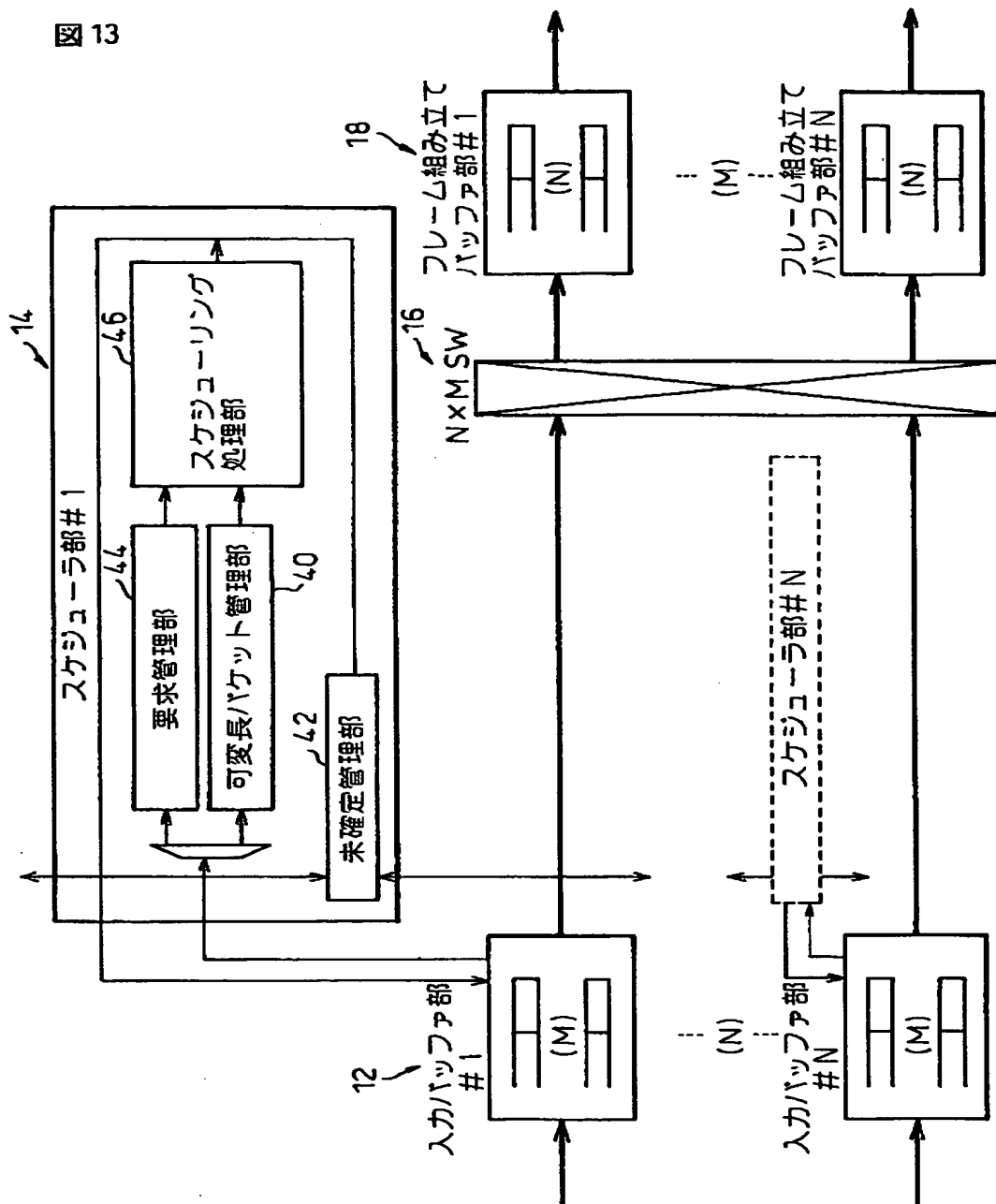


【図 12】



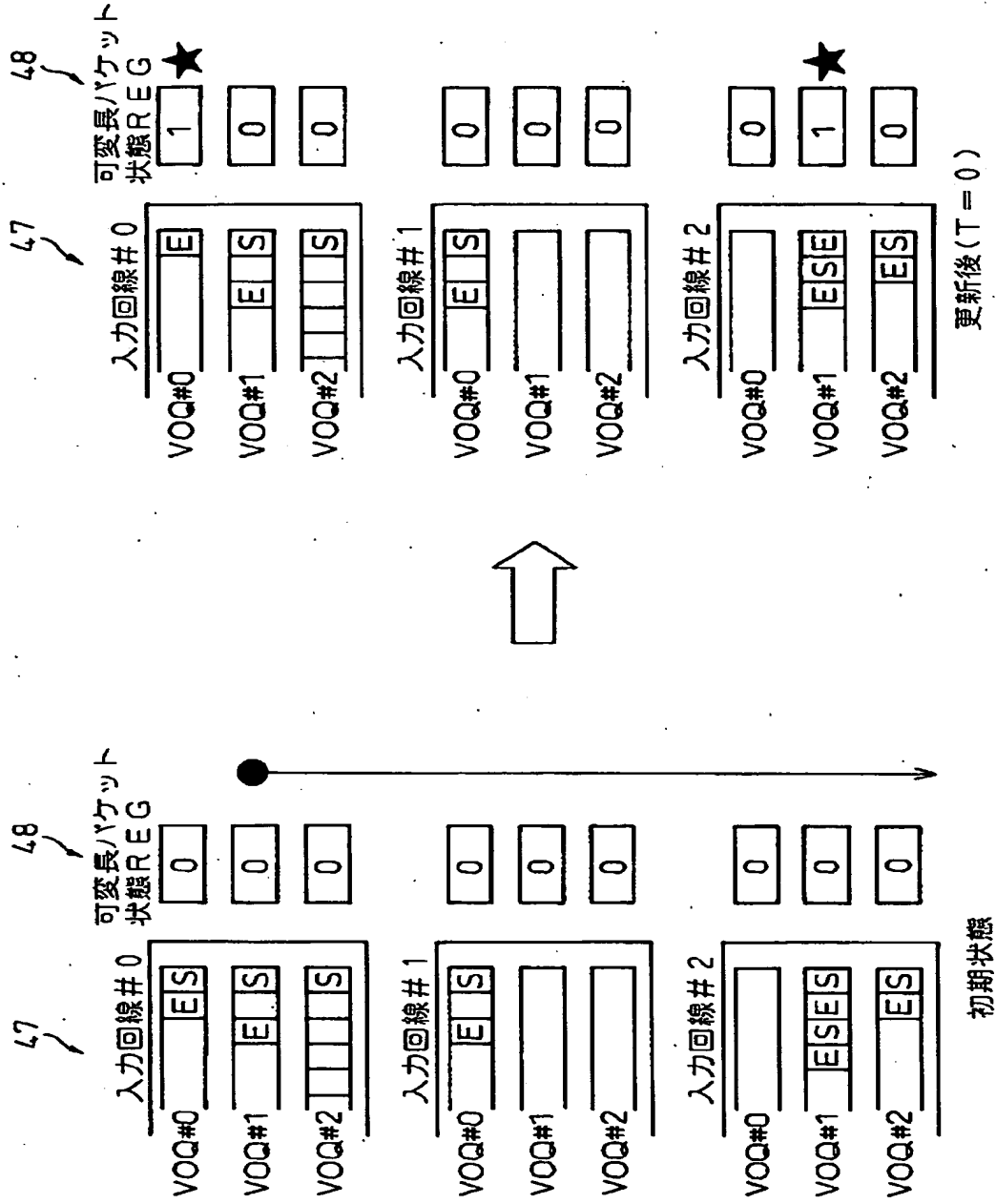
【図 13】

図 13

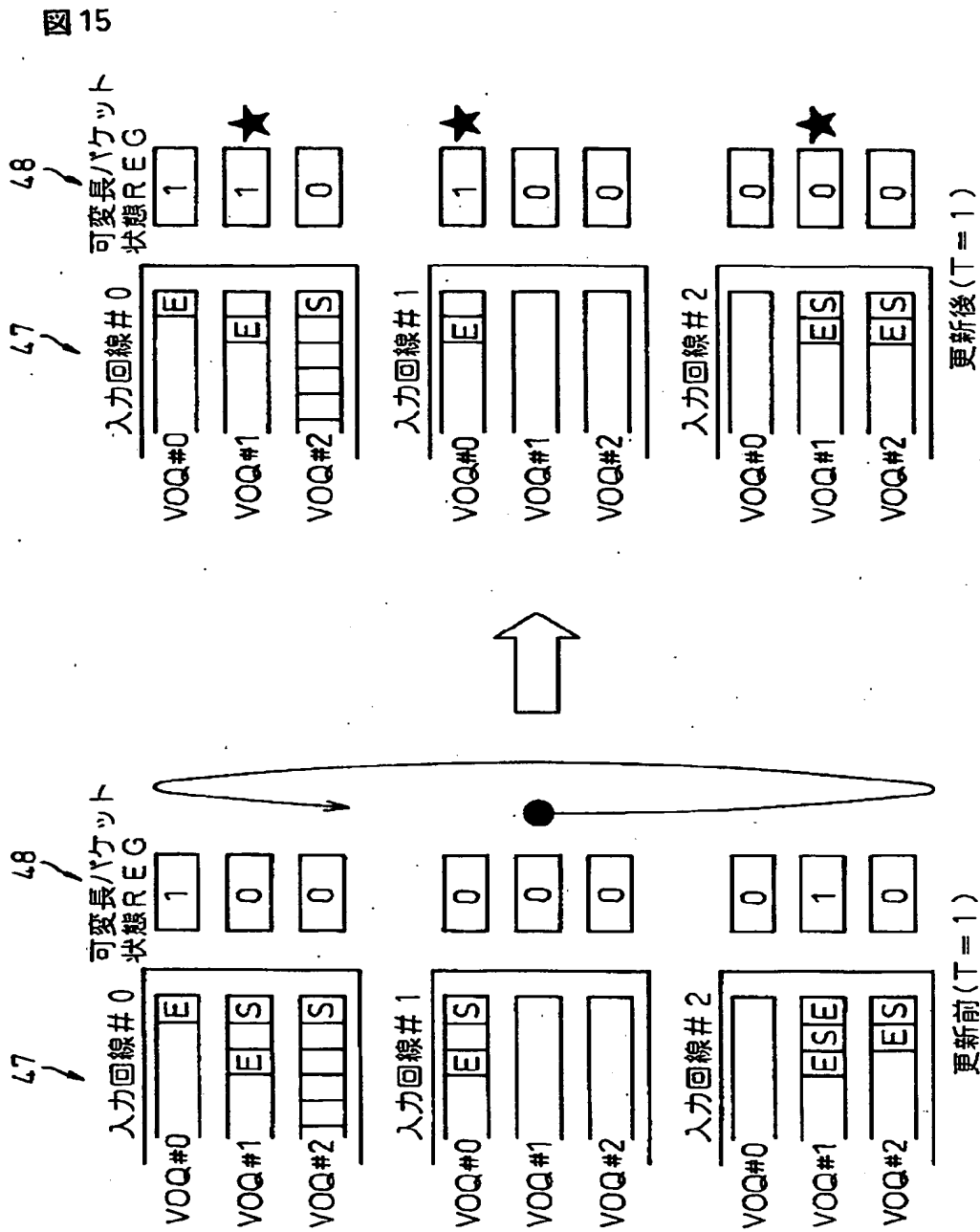


【図 1 4】

図 14

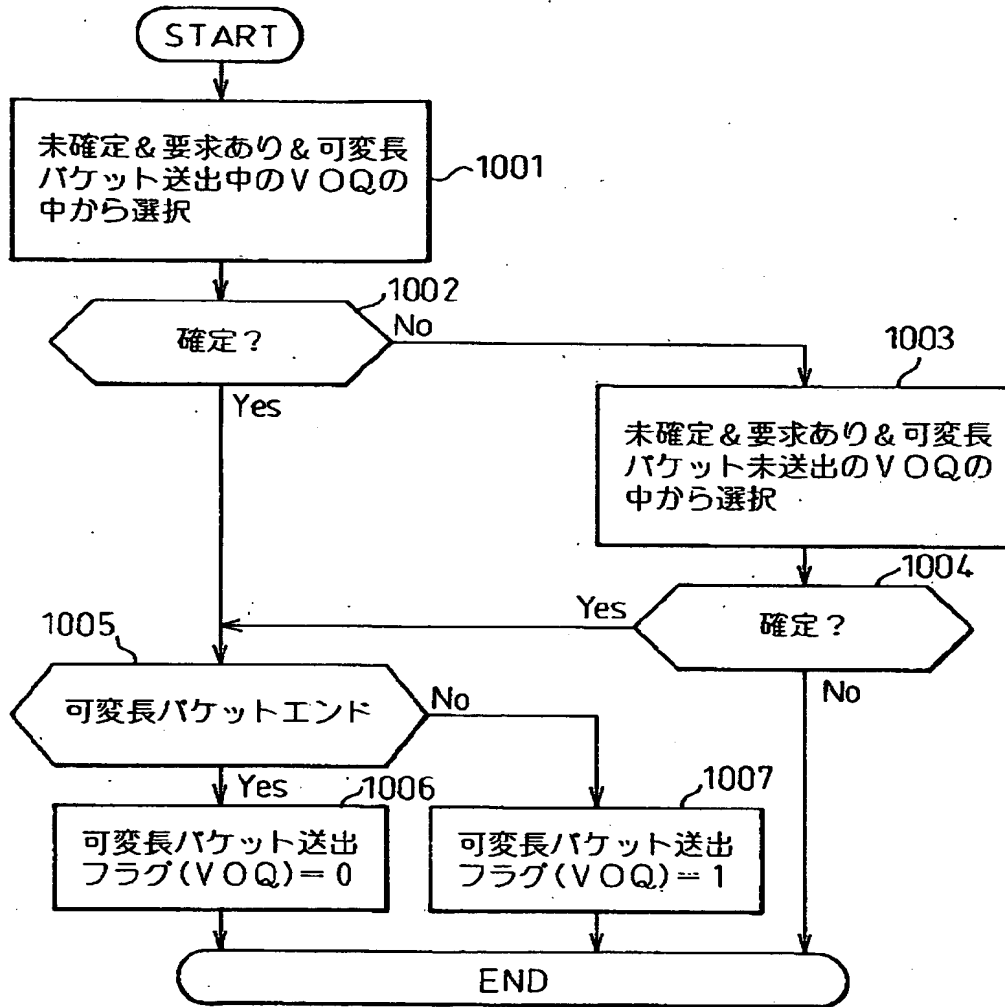


【図 15】

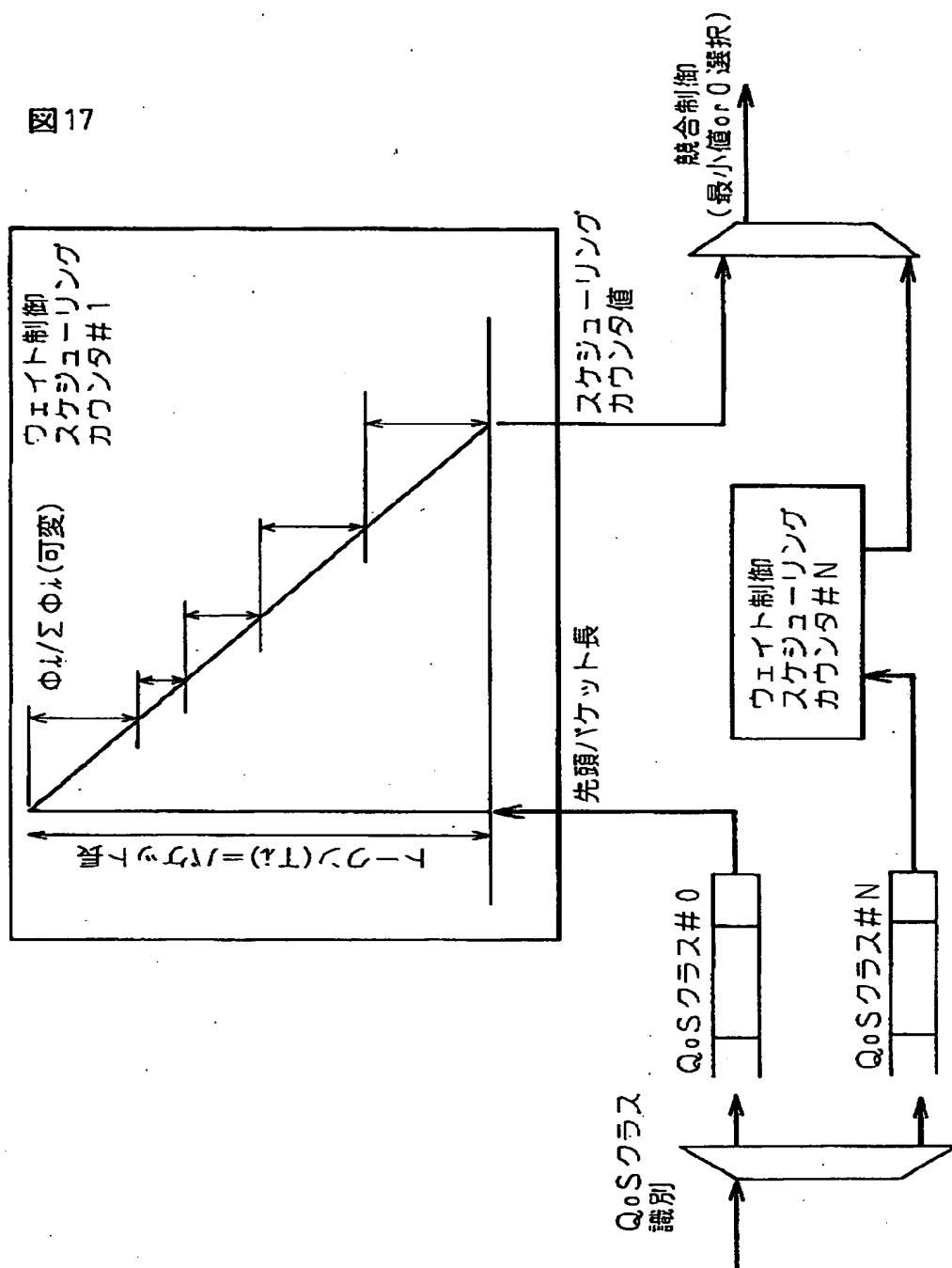


【図 1 6】

図 16

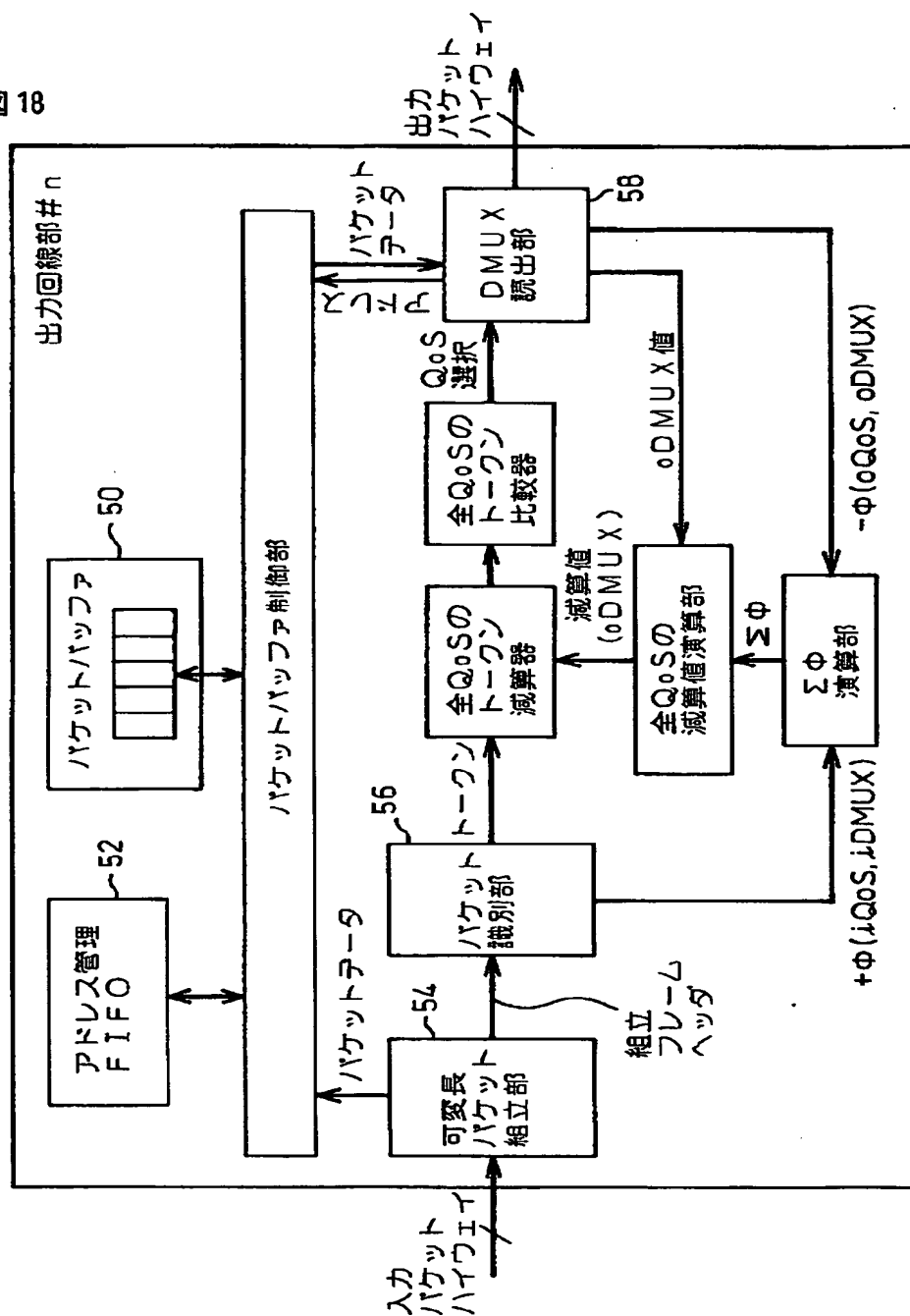


【図 17】



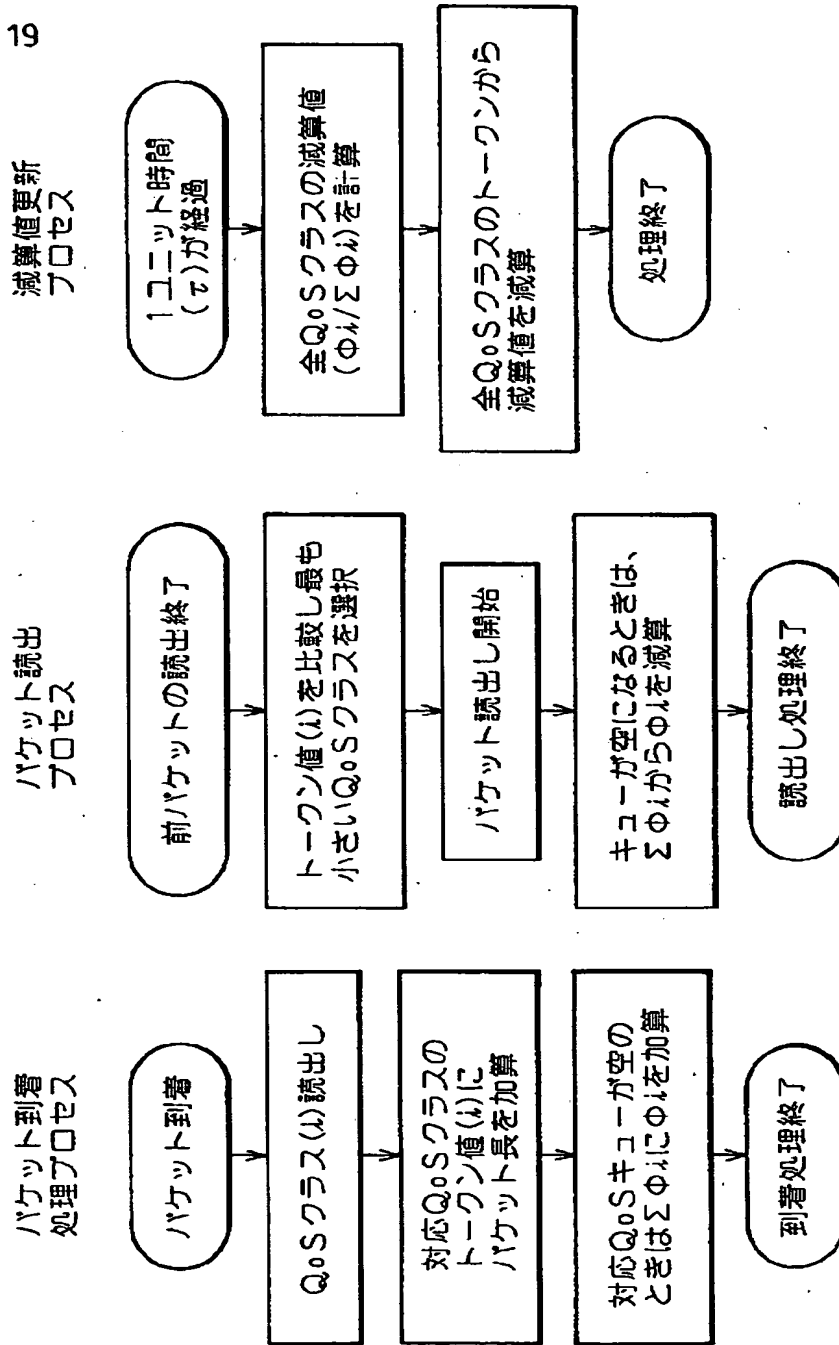
【图 18】

图 18

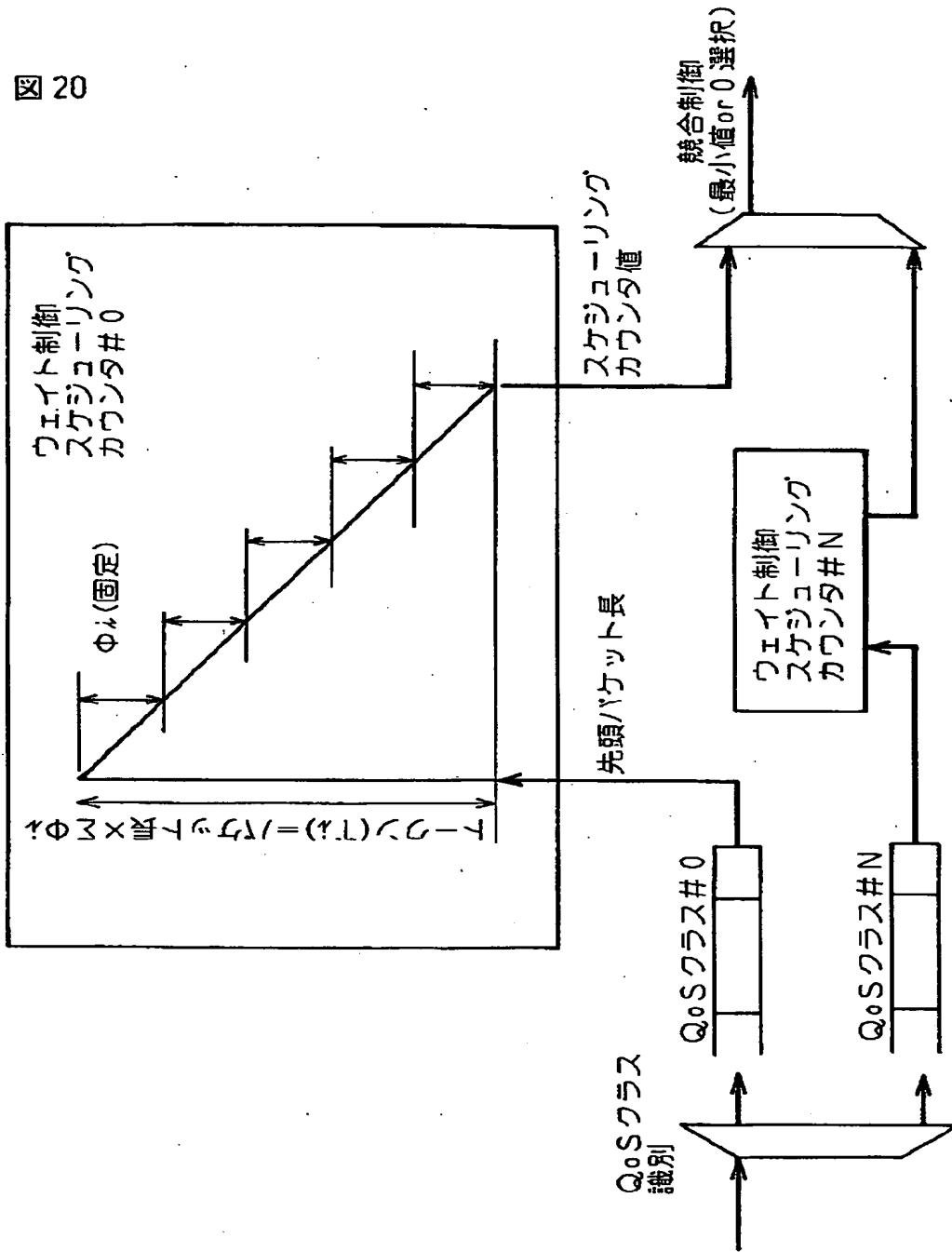


【図 19】

図 19

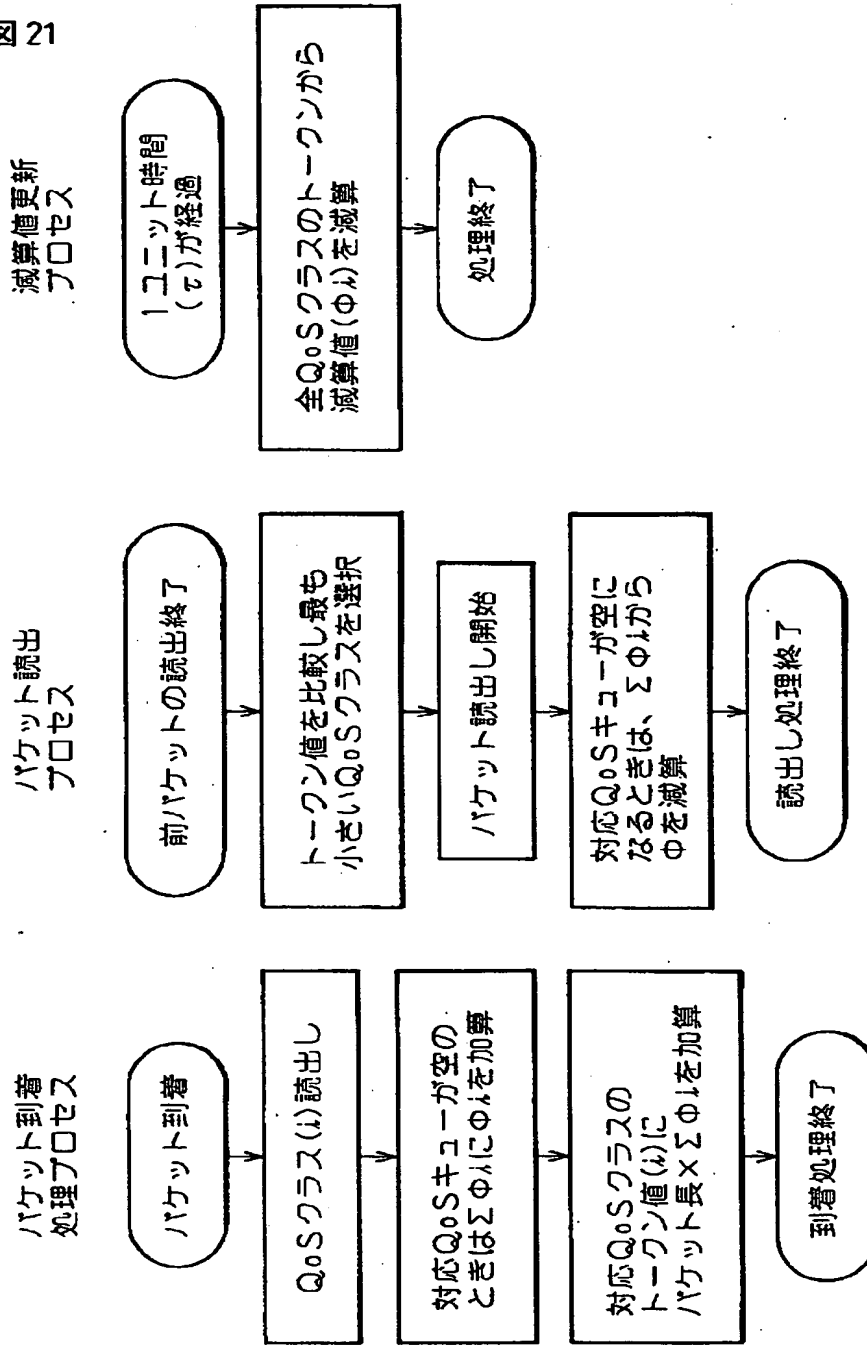


【図 20】



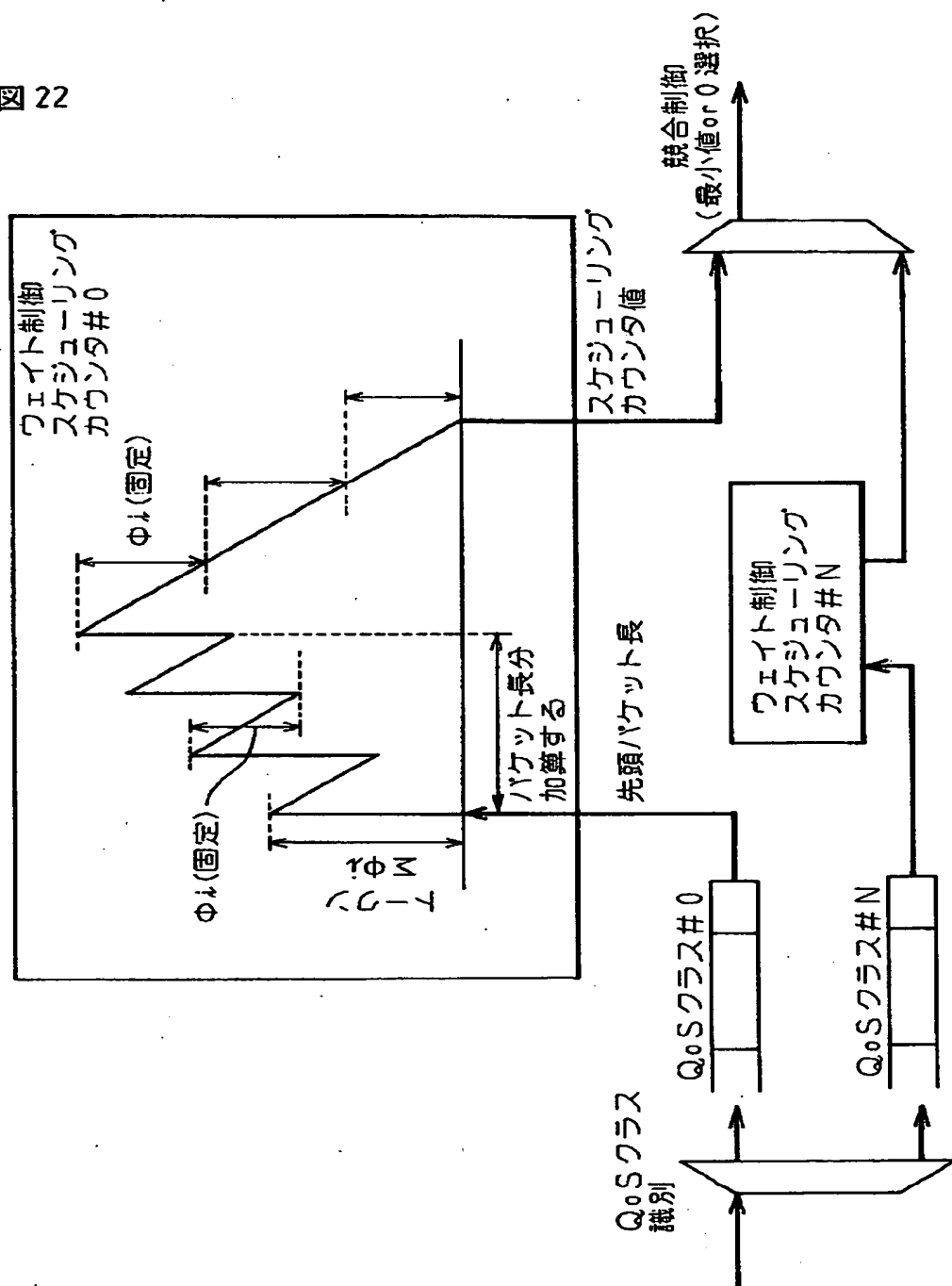
【図 21】

図 21

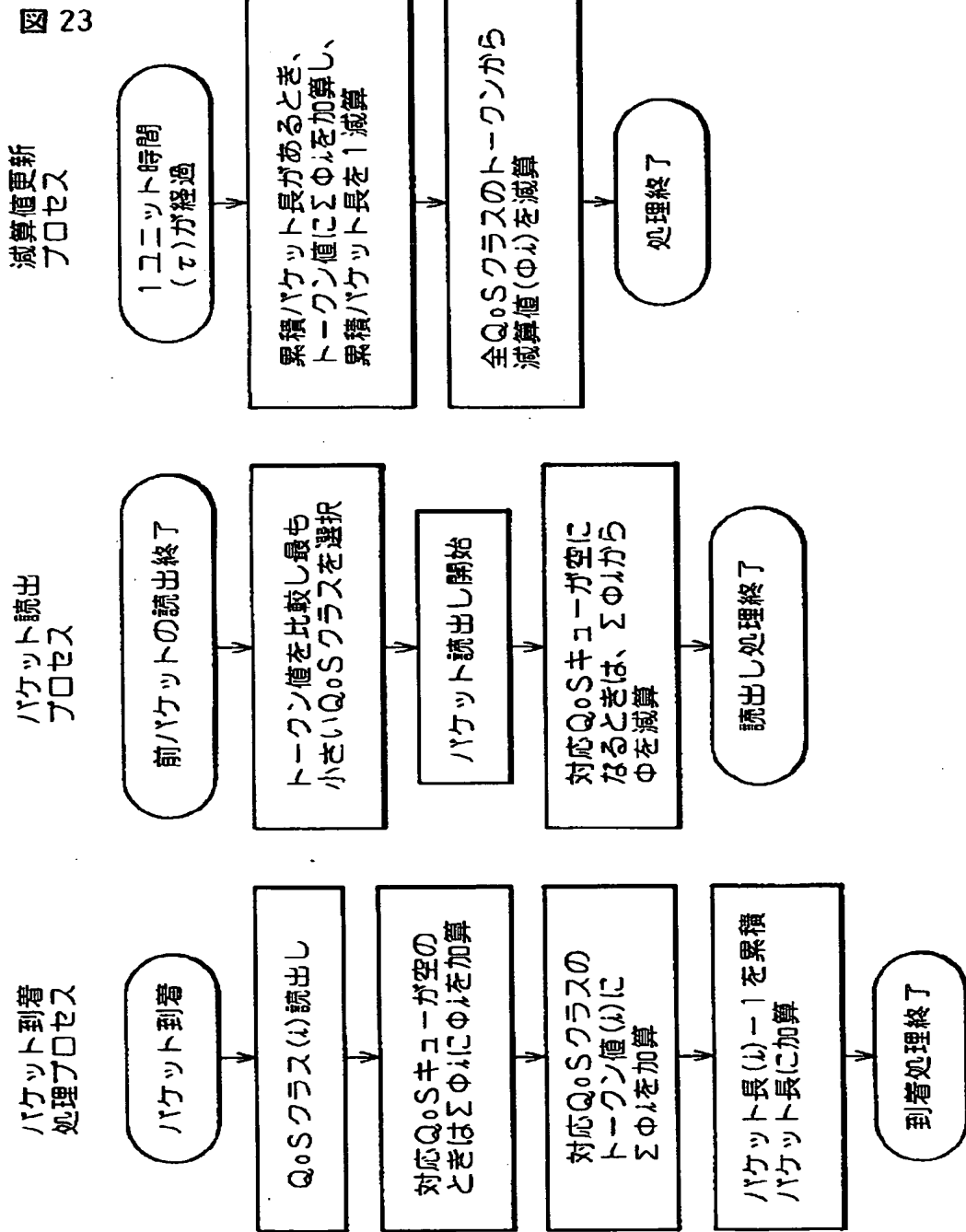


【图 2 2】

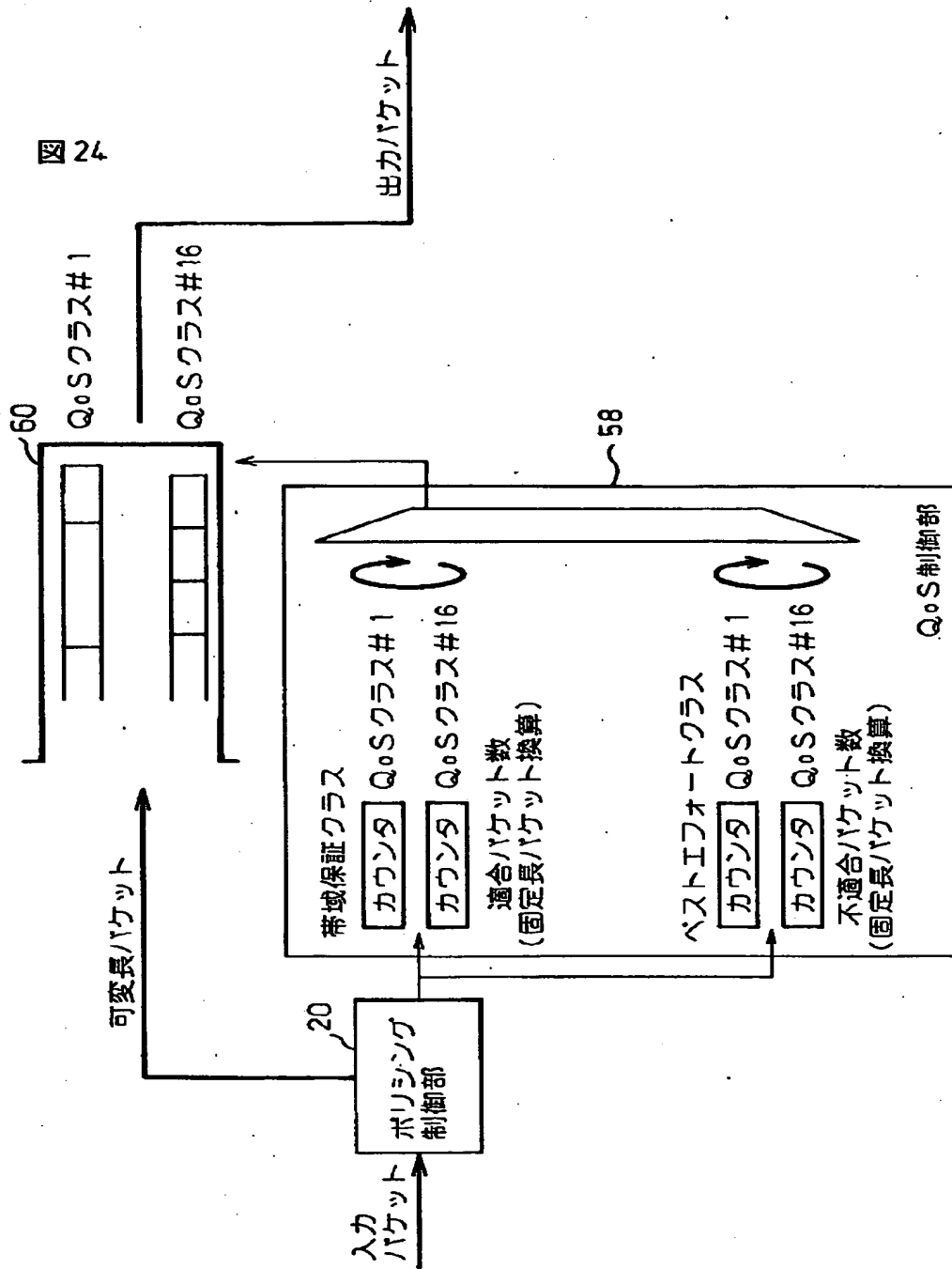
图 22



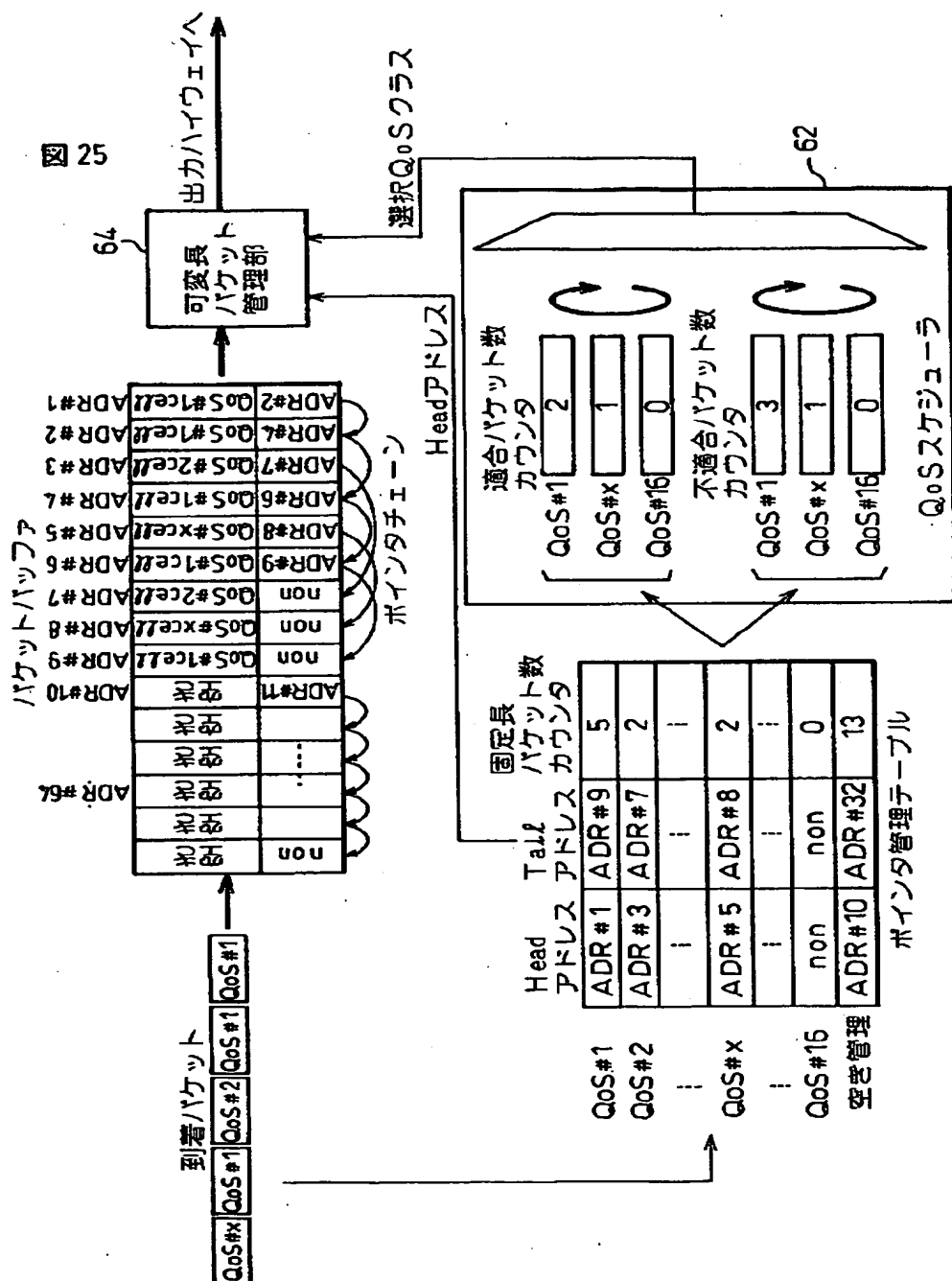
【図23】



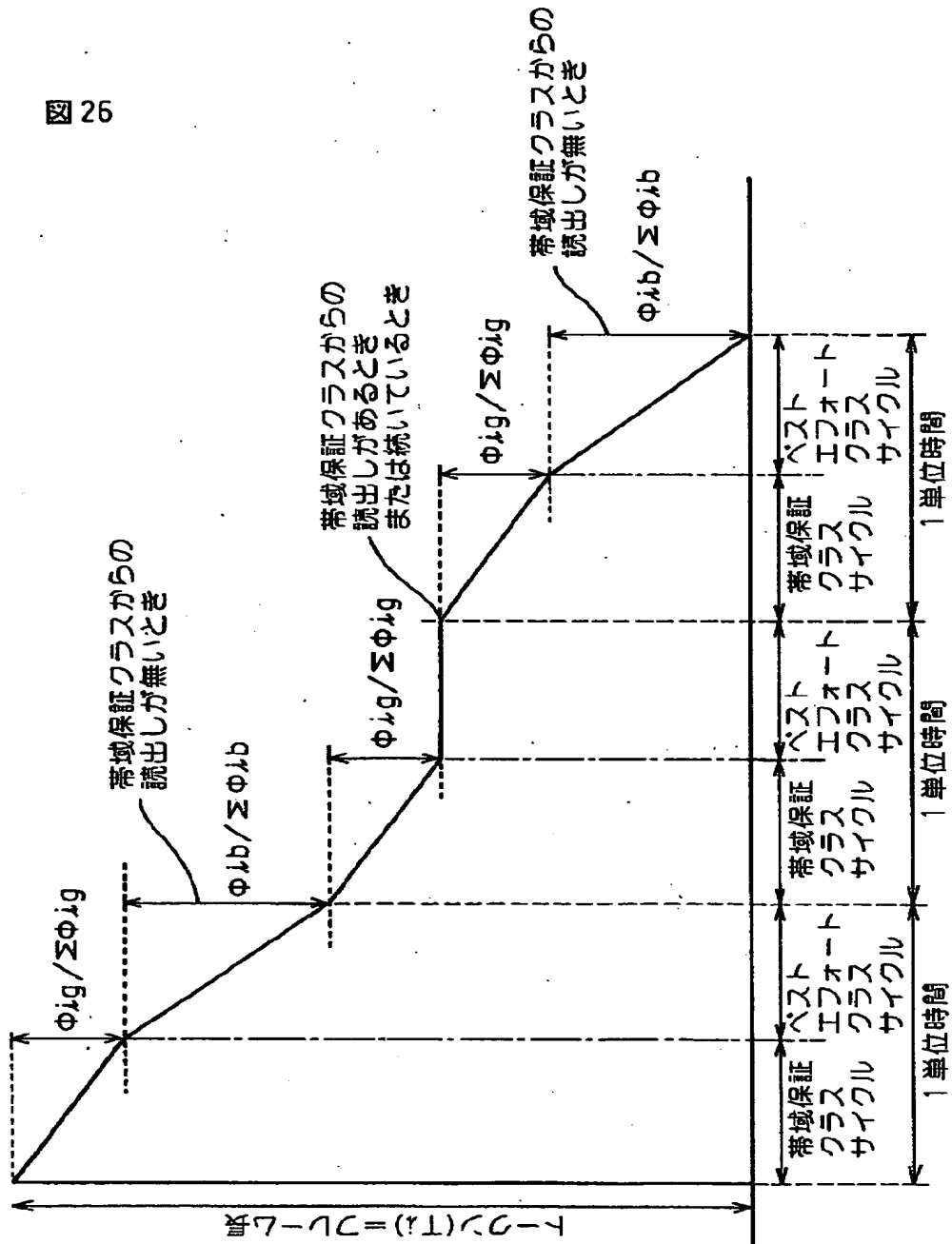
【図 24】



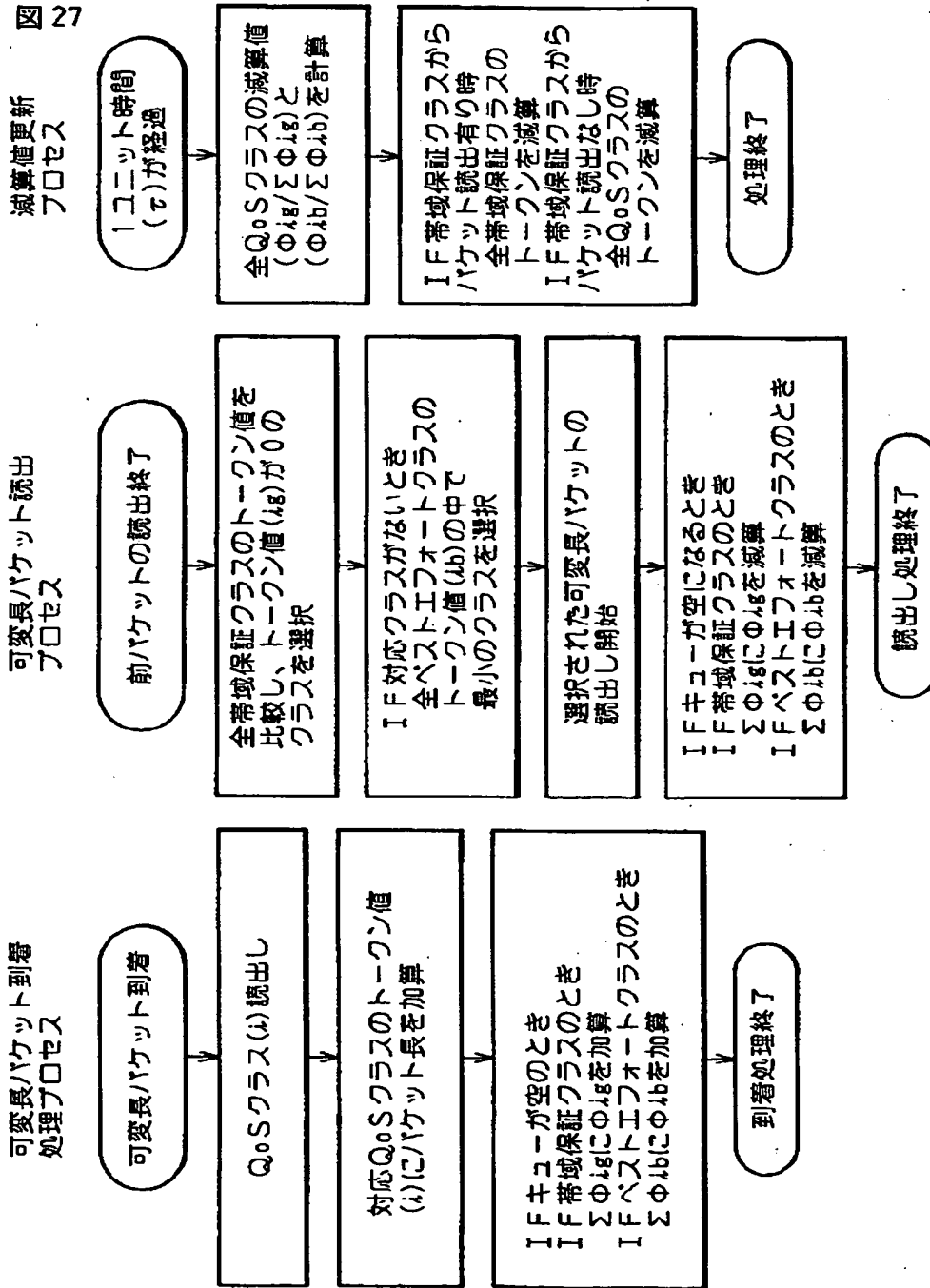
【図 25】



【図 2.6】

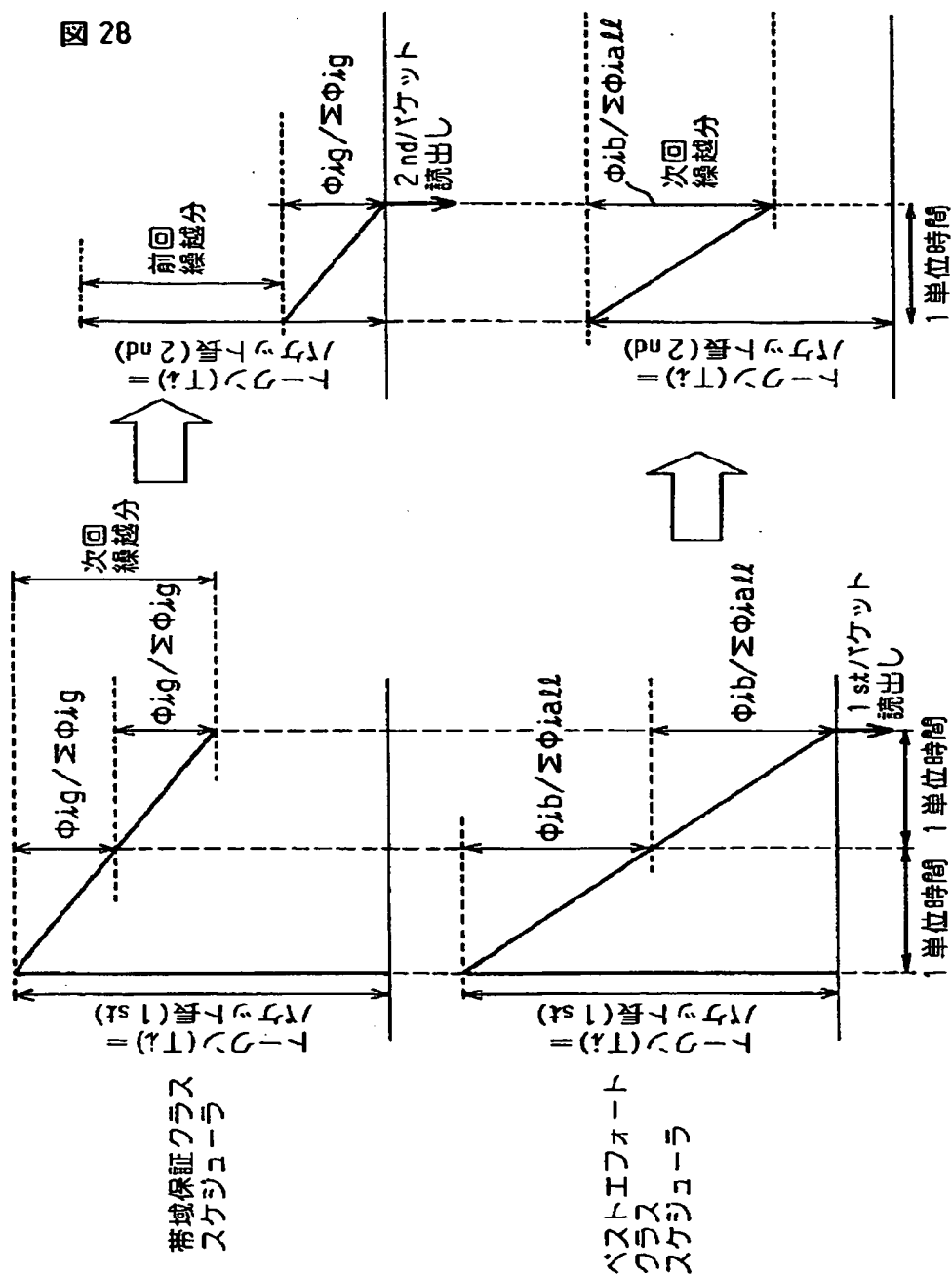


【図 27】

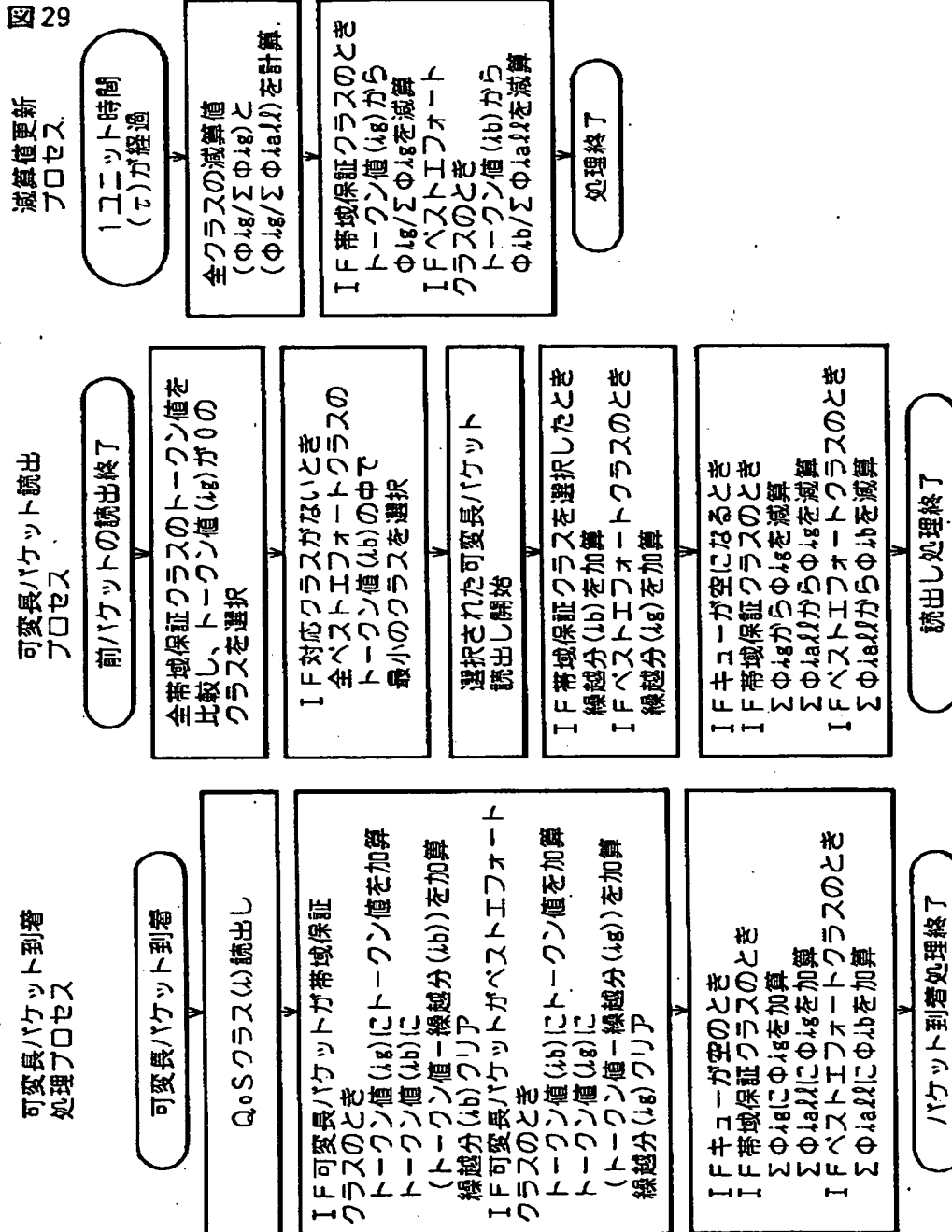


【図 28】

図 28

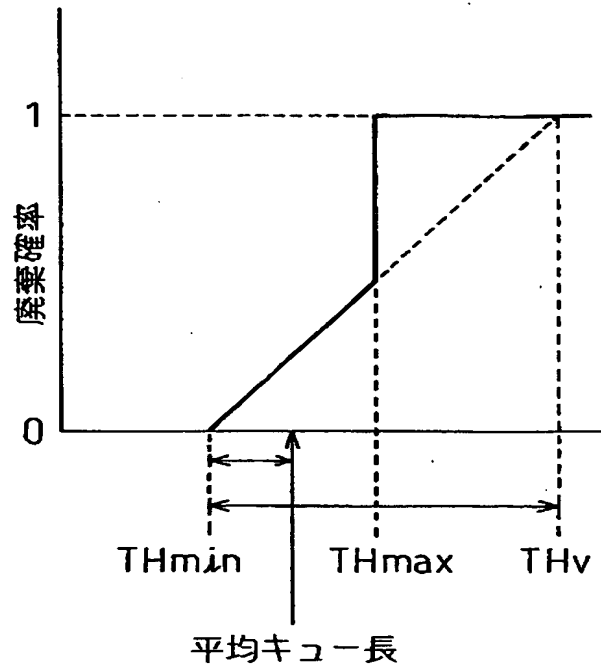


【図 2 9】



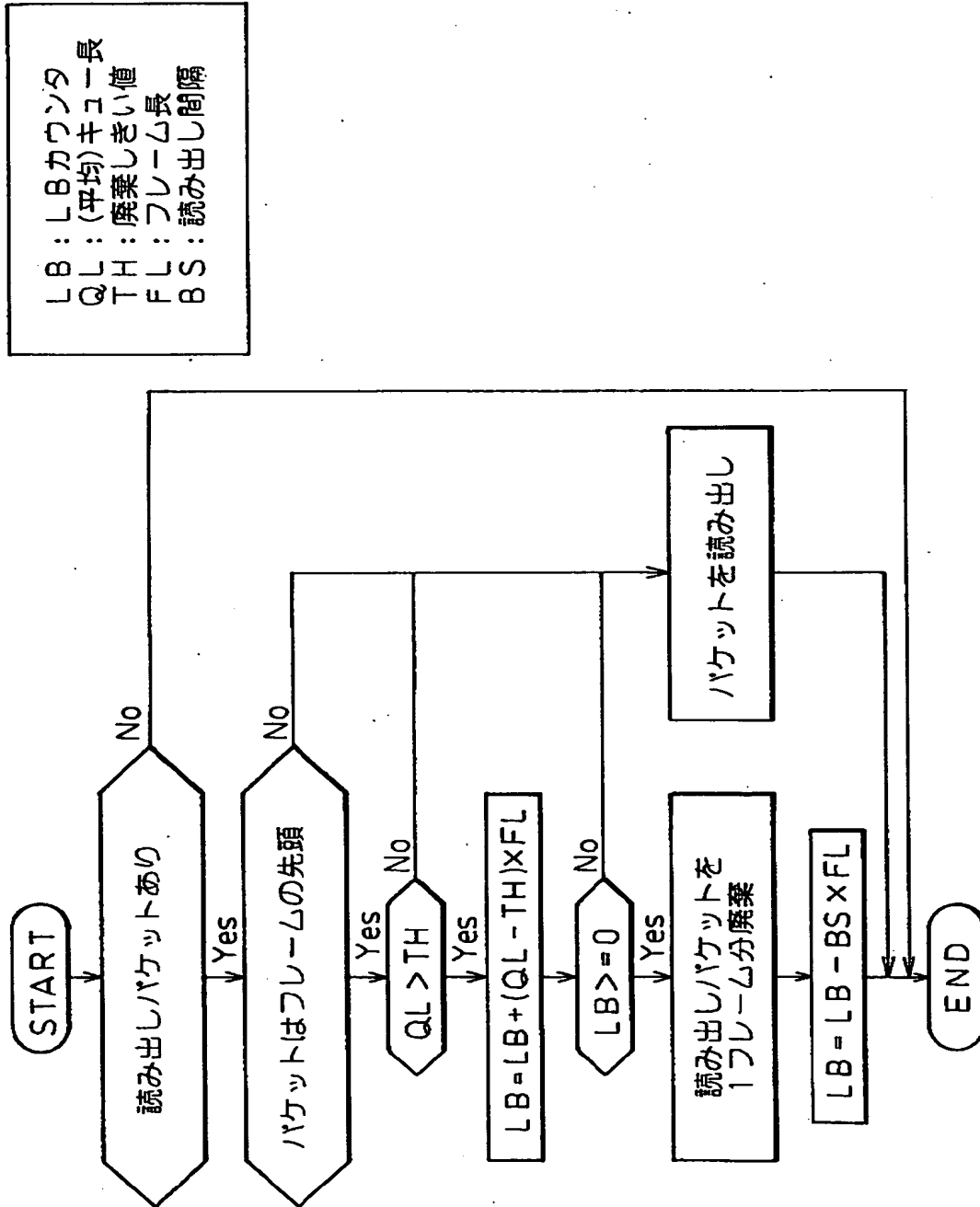
【図 3 0】

図 30



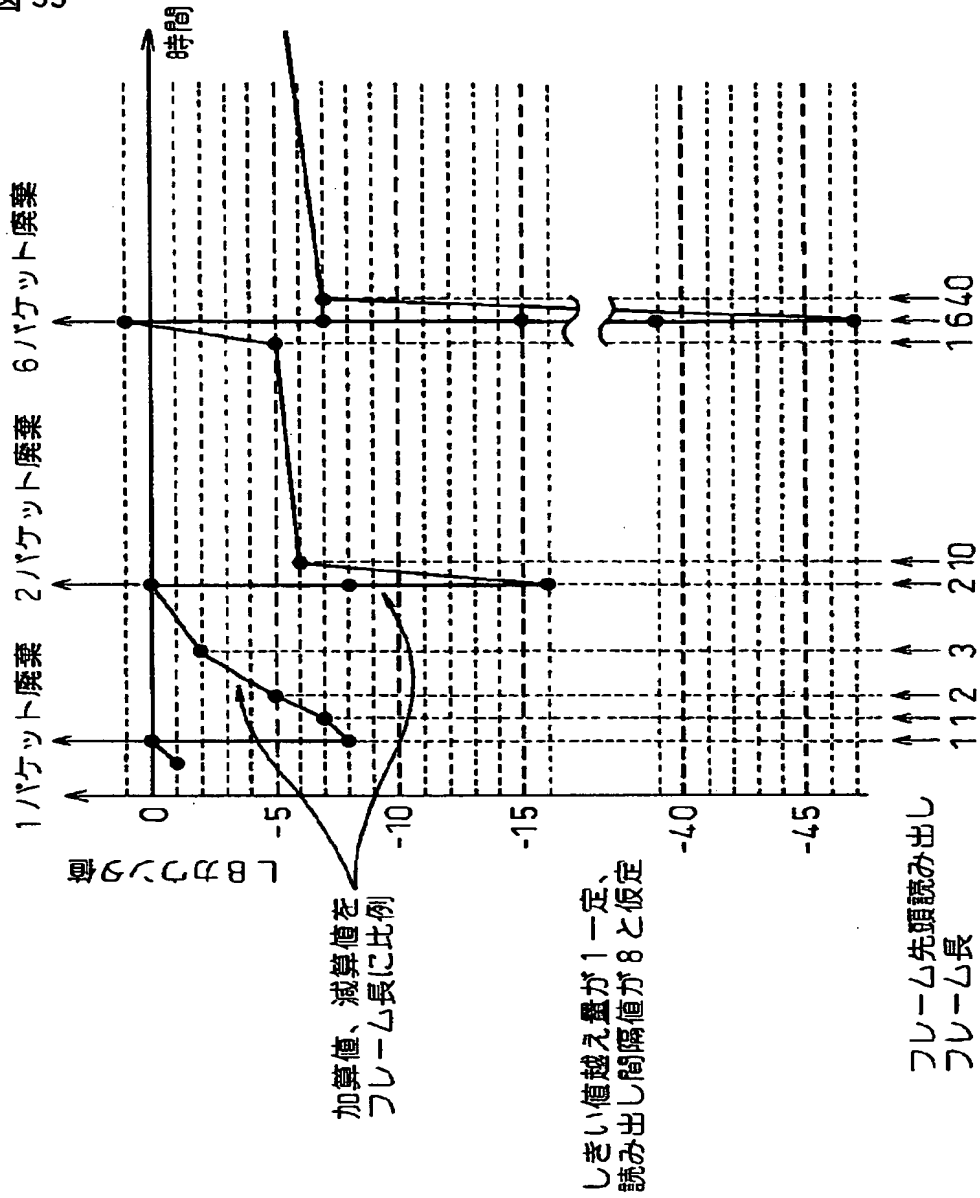
【図 32】

図 32



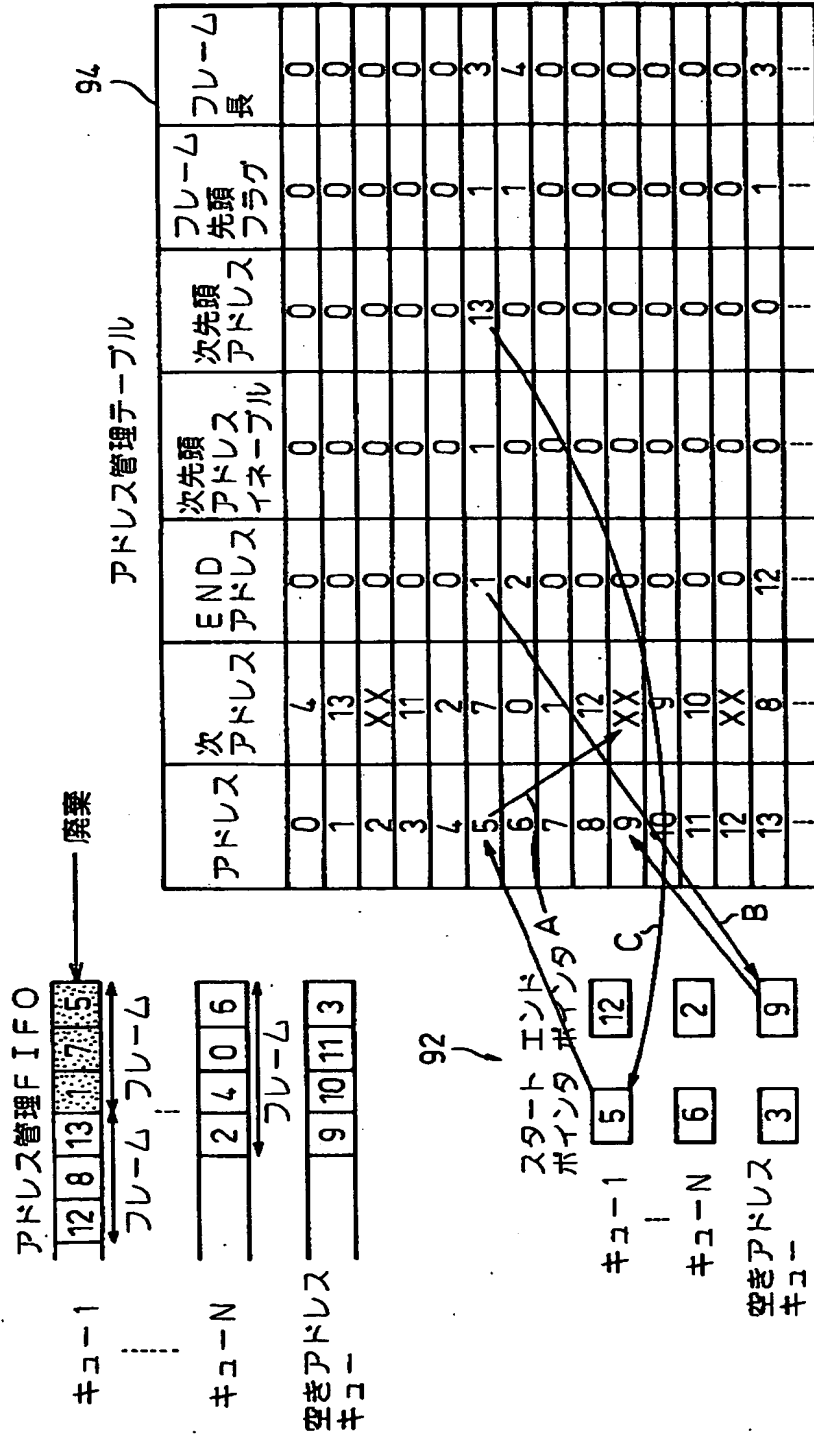
【図 3 3】

圖 33



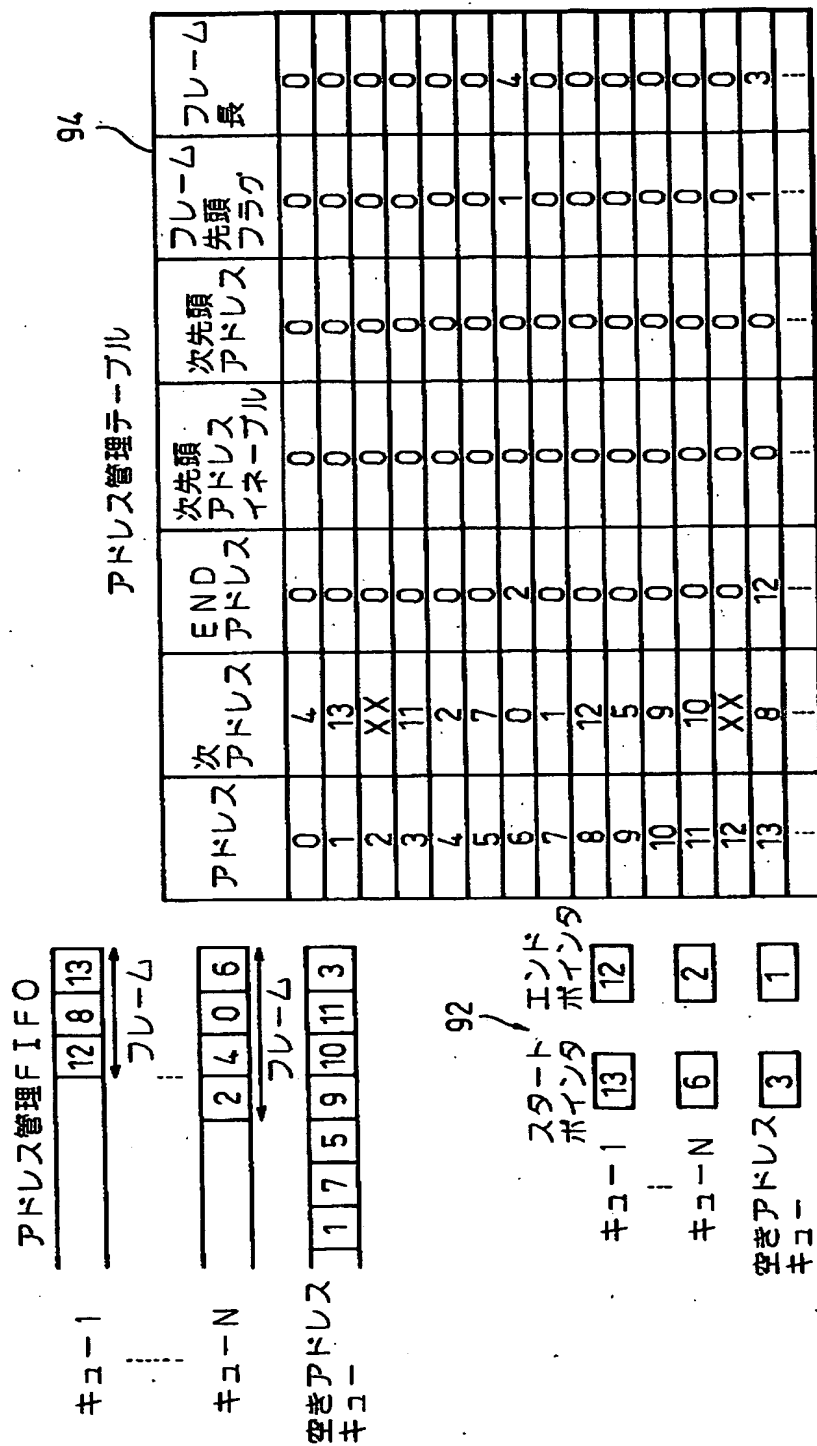
【図34】

図 34

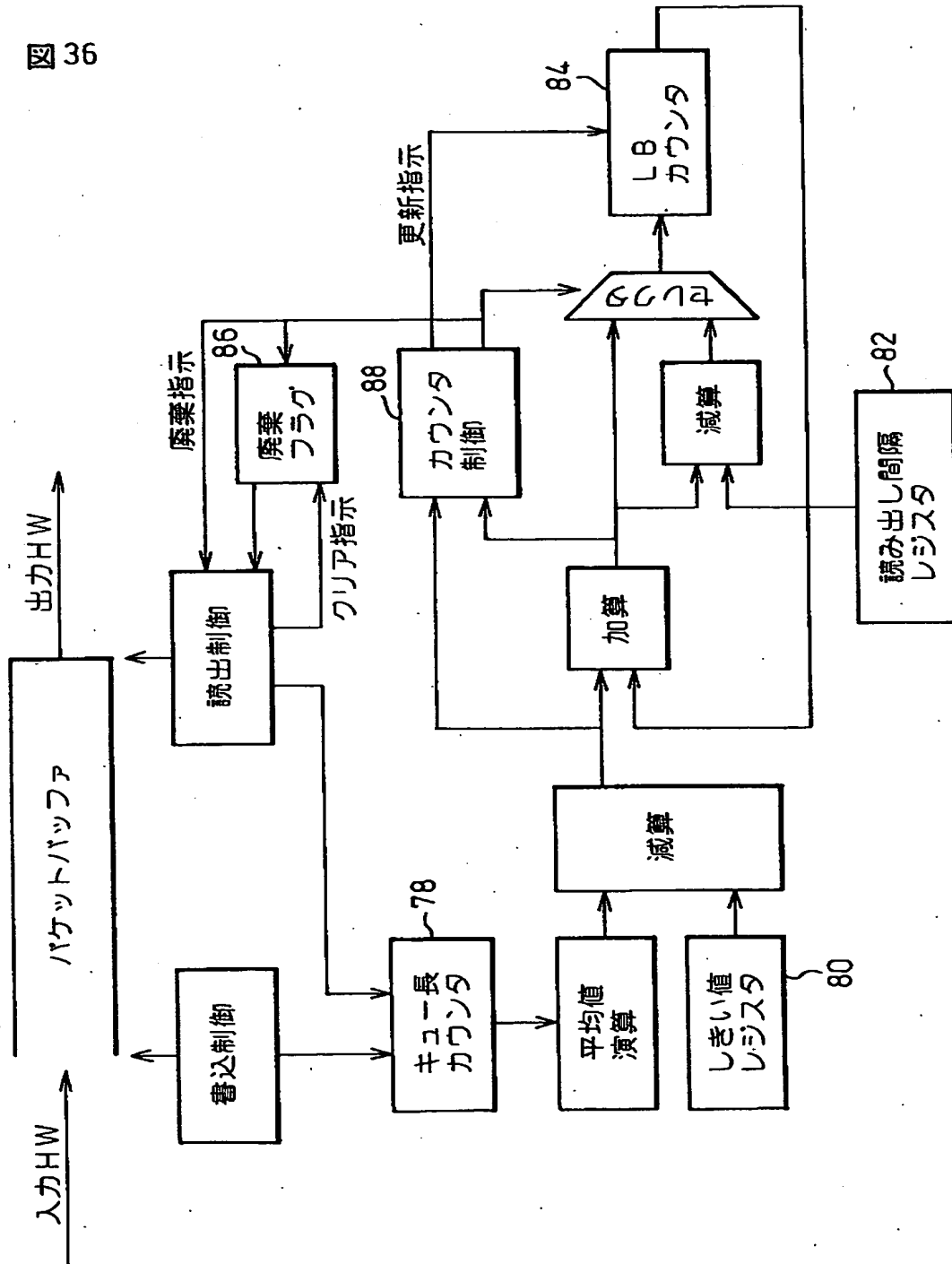


【図35】

図35



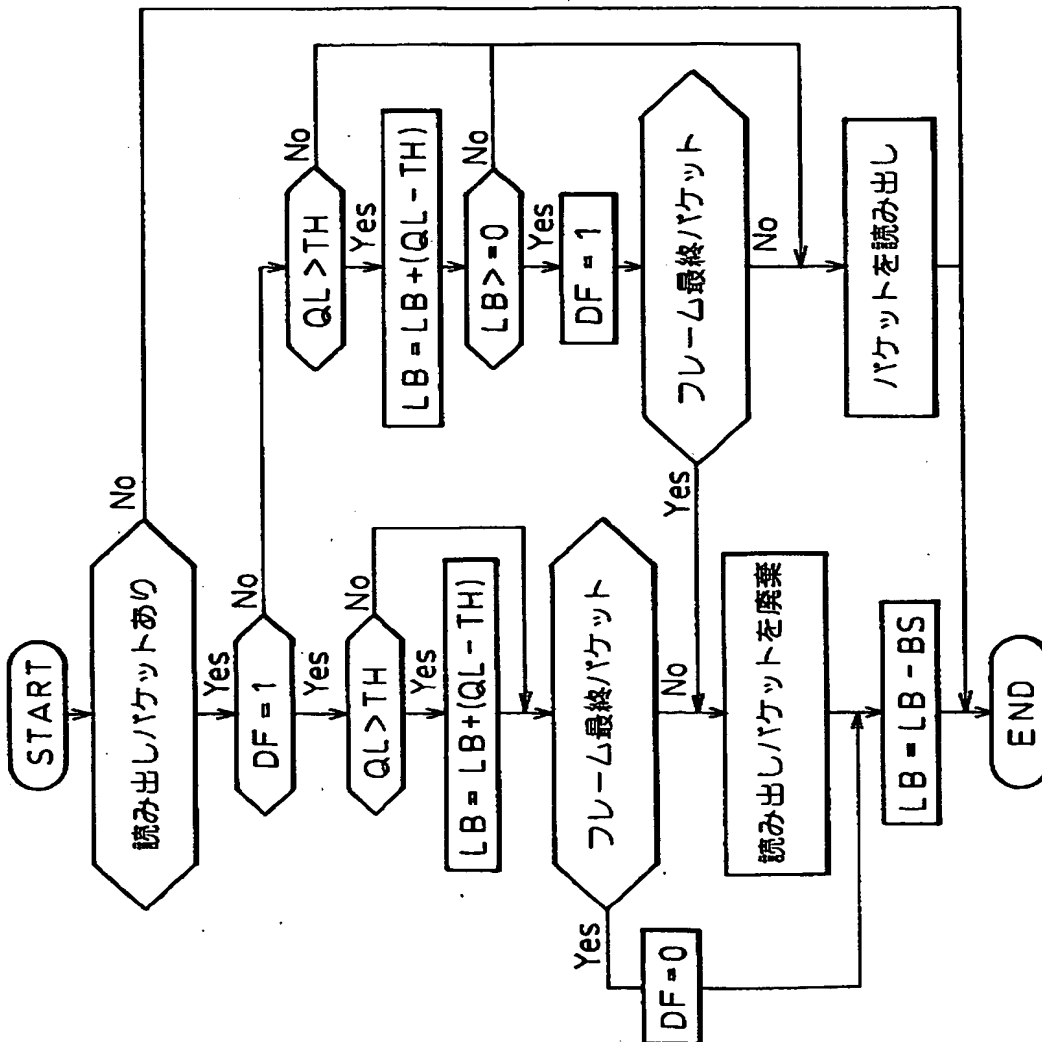
【図 36】



【図 37】

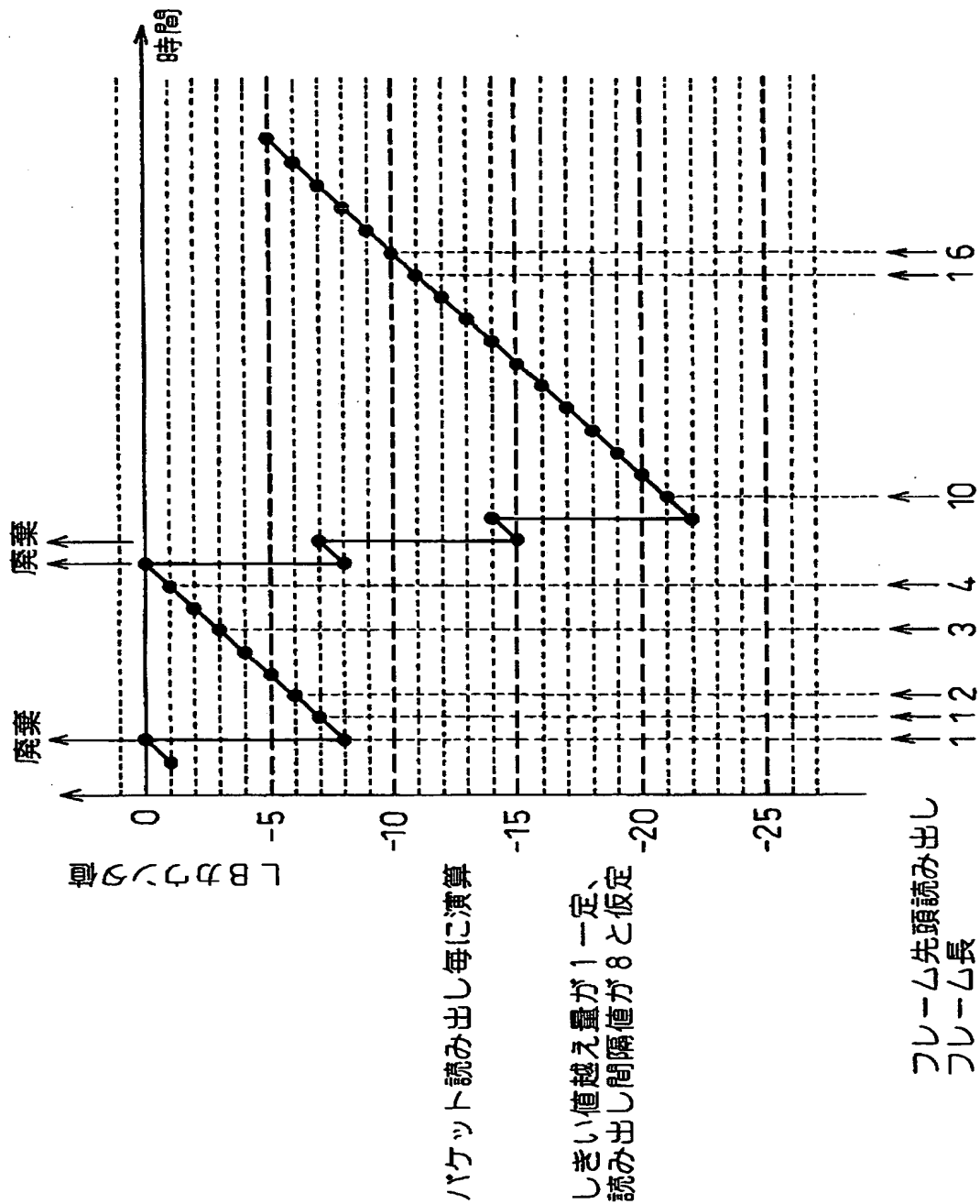
図 37

LB: LBカウンタ
QL: (平均)キュー長
TH: 廃棄しきい値
DF: 廃棄フラグ
BS: 読み出し間隔



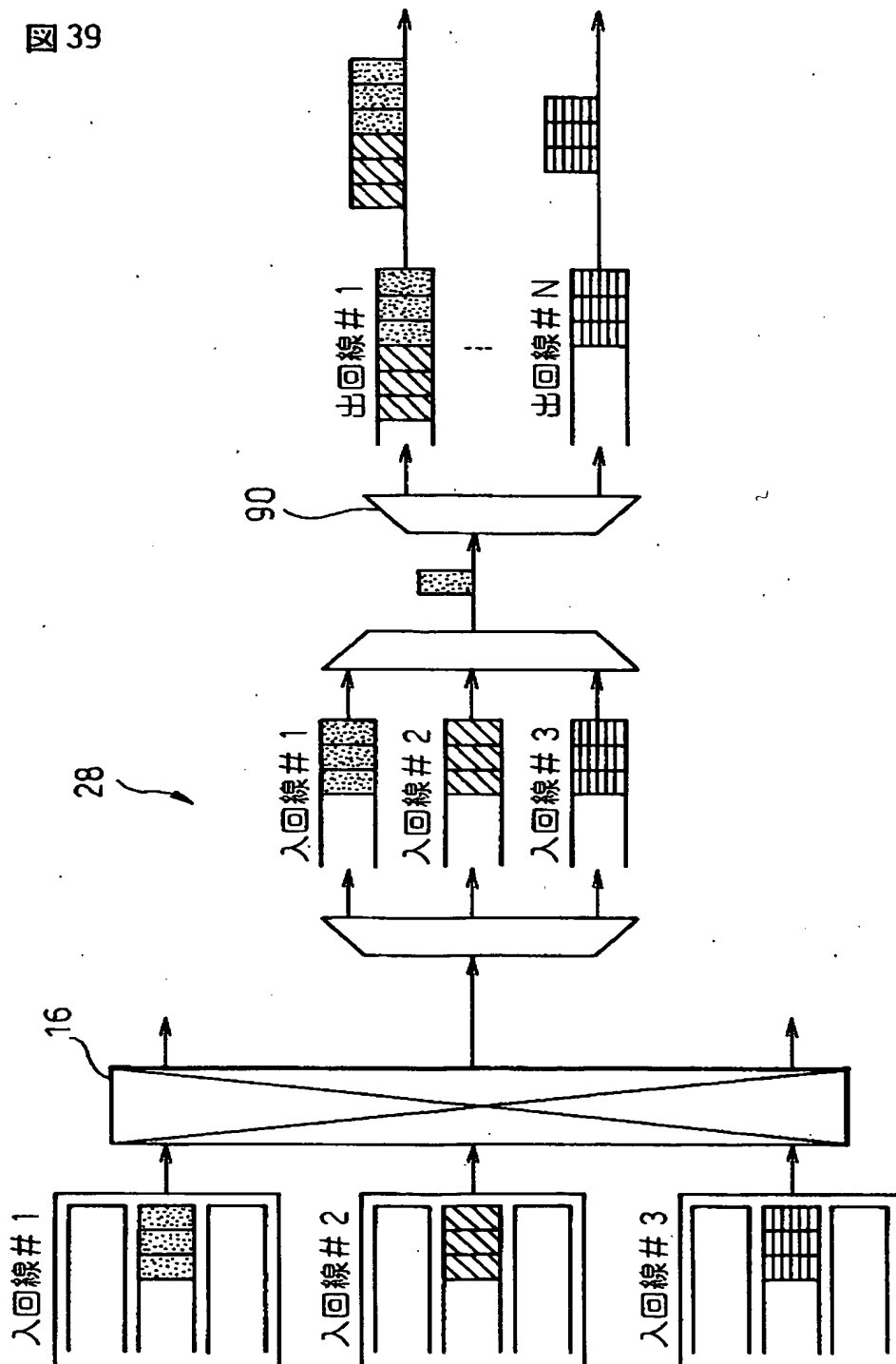
【図38】

図38

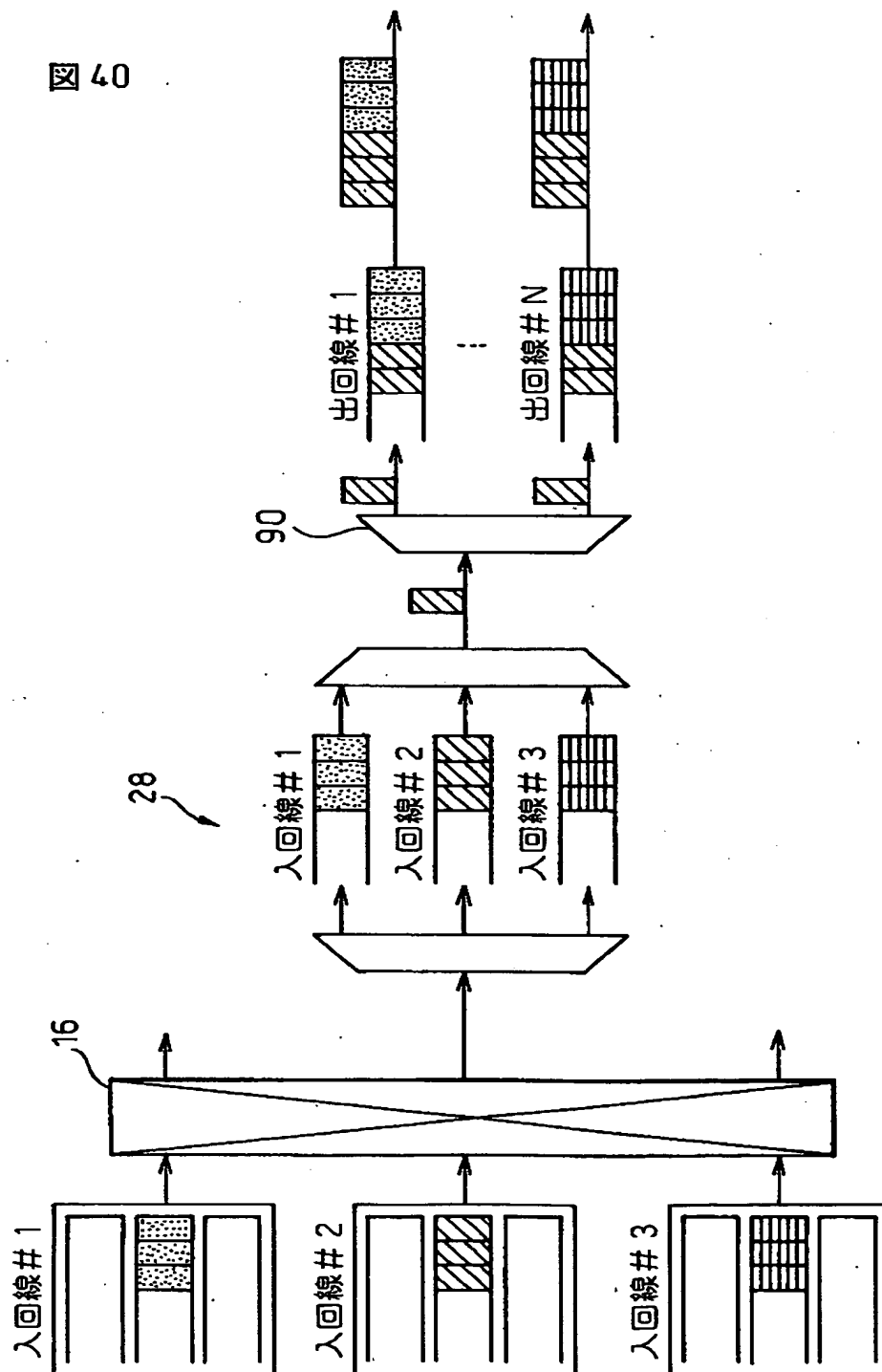


【図 39】

図 39

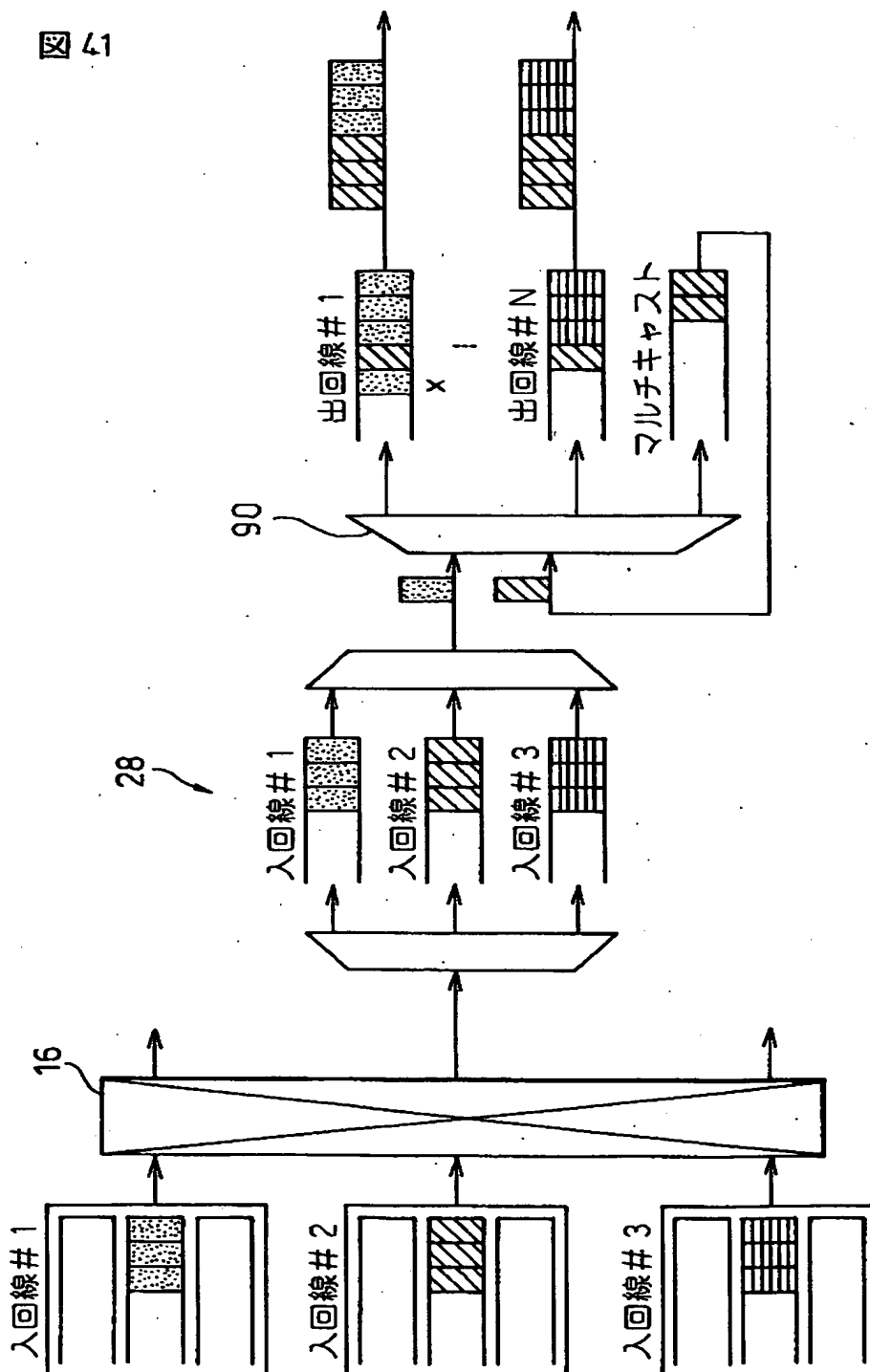


【図 40】

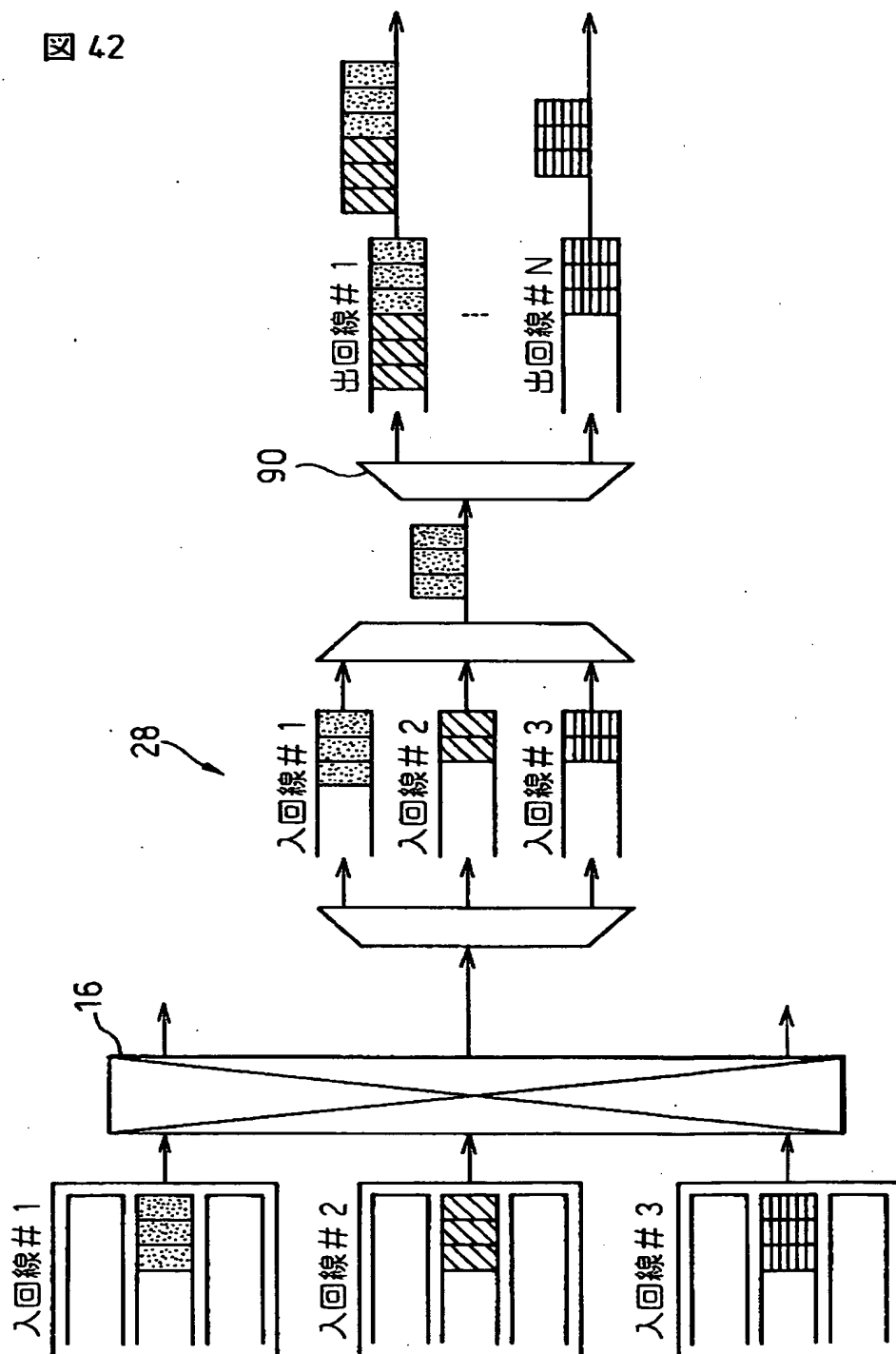


【図 41】

図 41

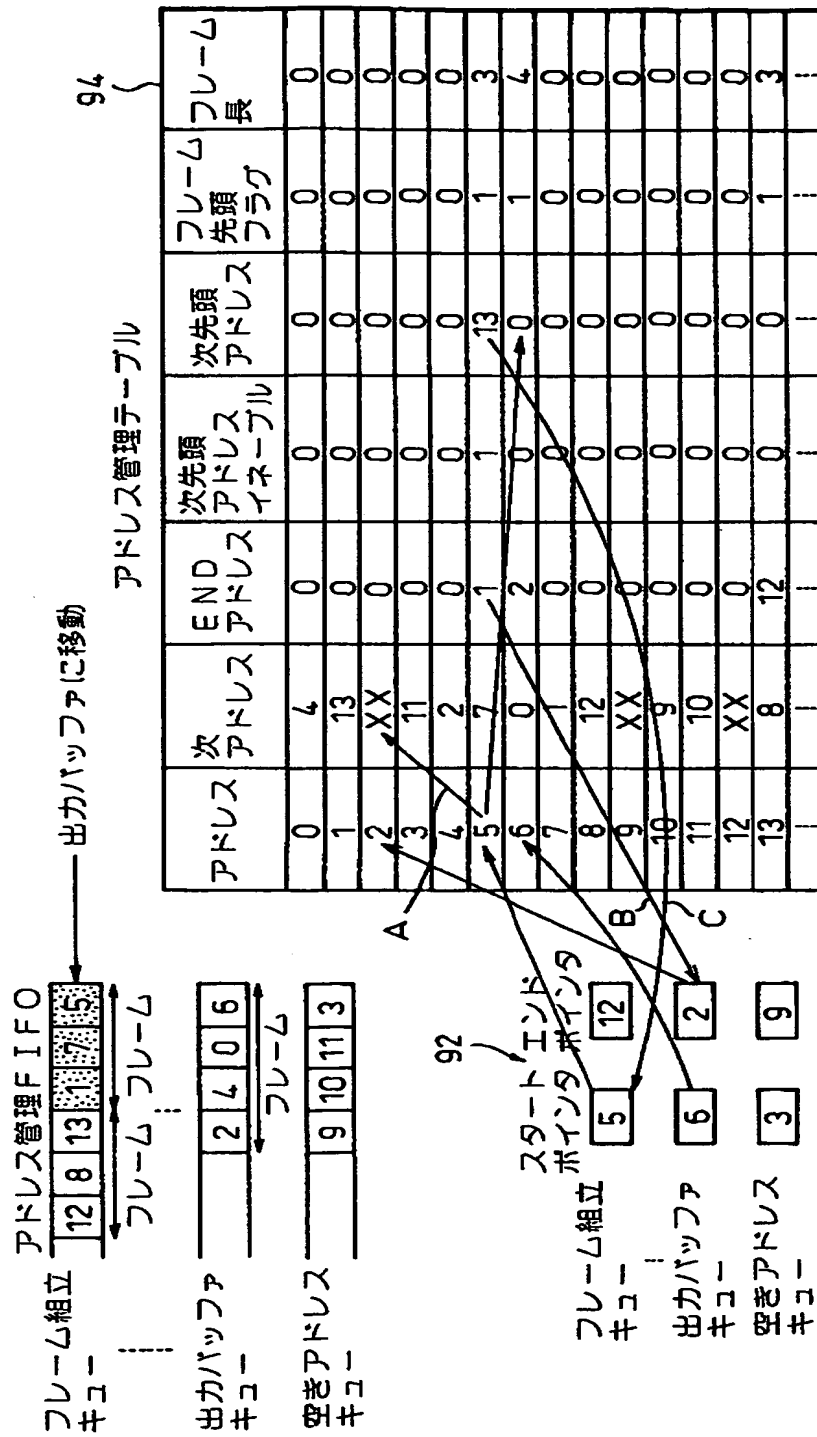


【図 42】



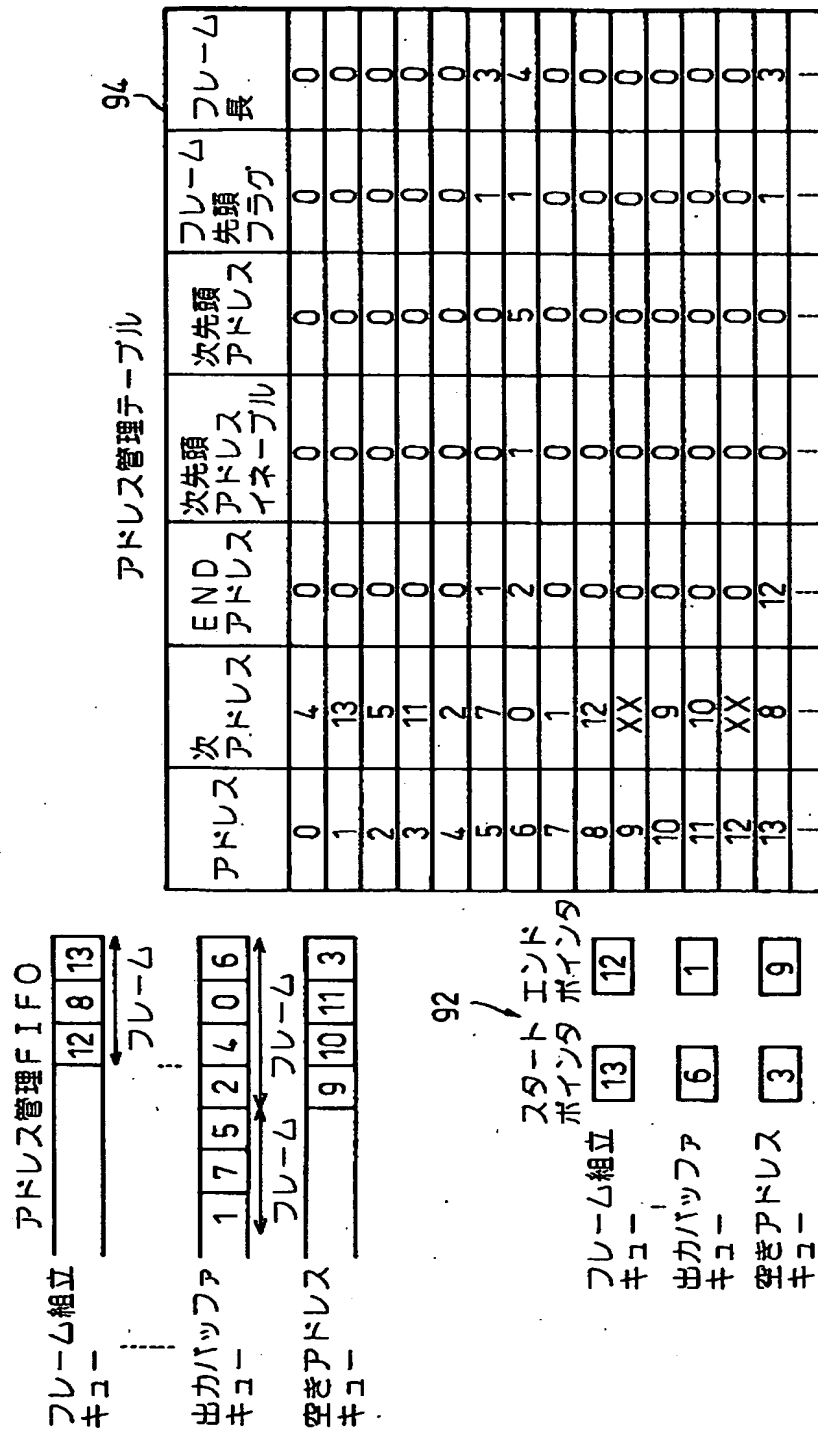
【図 43】

図 43

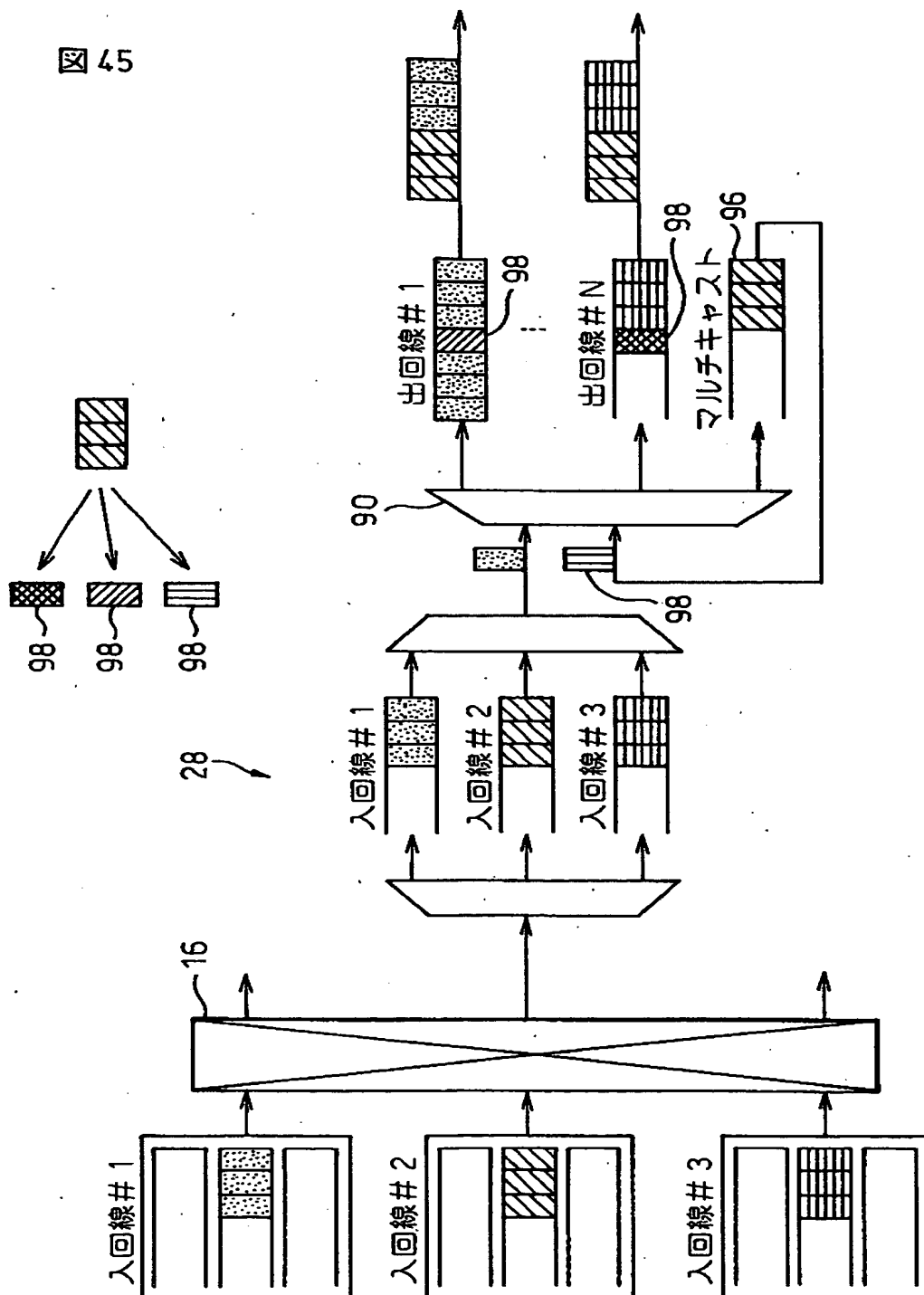


【図 4 4】

図 44

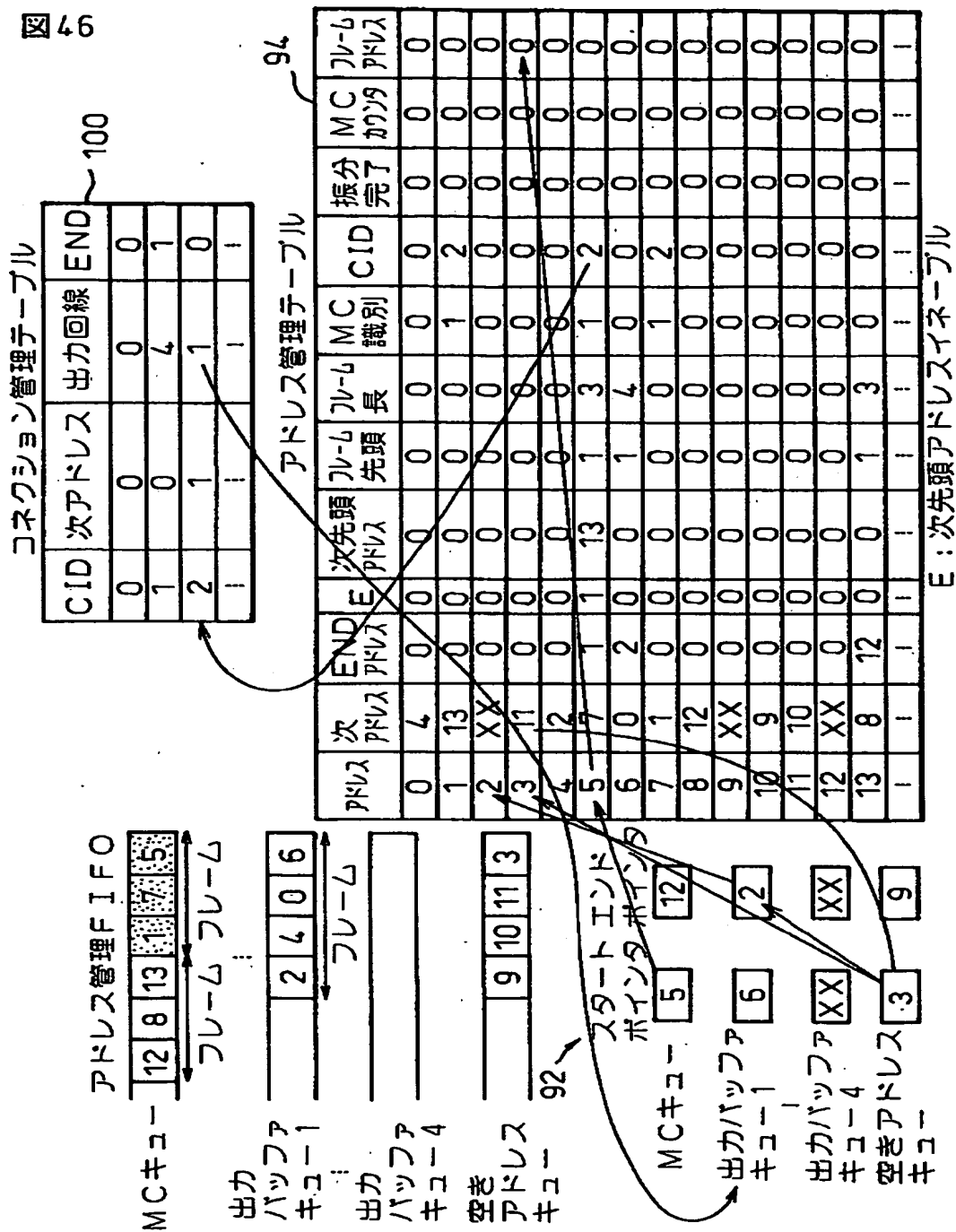


【図 45】



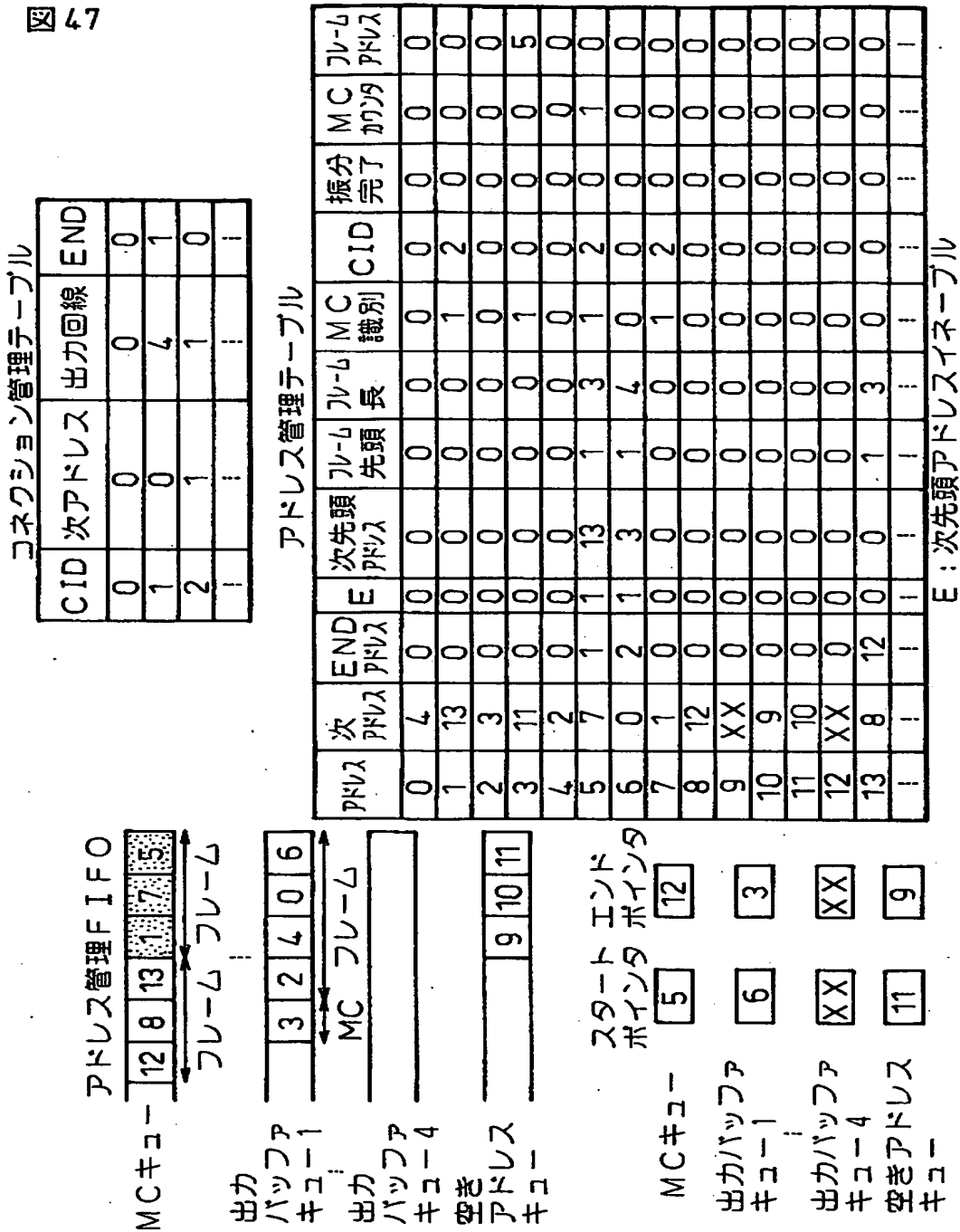
【図 4 6】

図 46

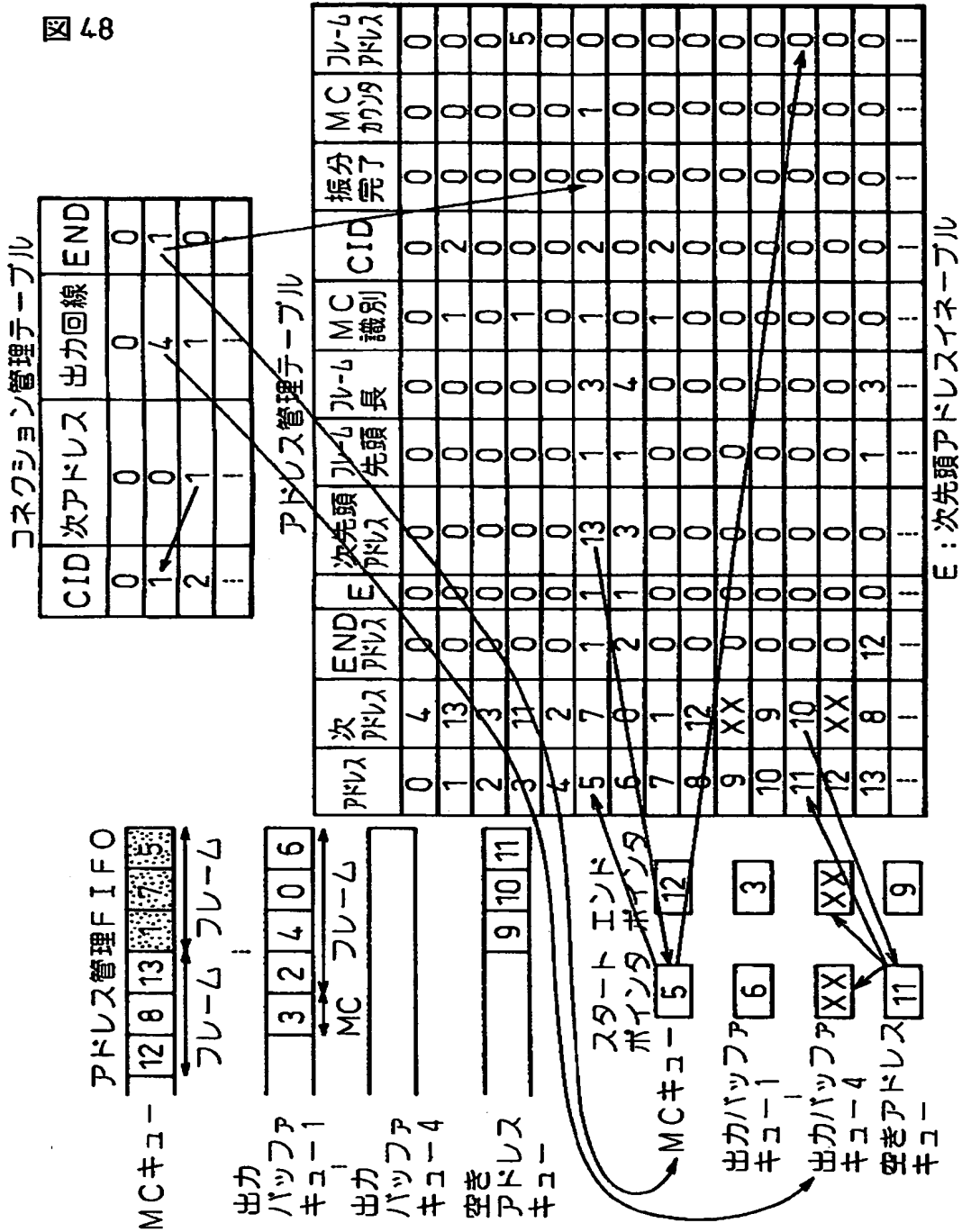


【図47】

図47



【図 48】



【図 49】

図 49

コネクション管理テーブル

CID	次アドレス	出力回線	END
0	0	0	0
1	0	4	1
2	1	1	0
1	1	1	1

アドレス管理テーブル

アドレス	次アドレス	END アドレス	E	次先頭 アドレス	フレーム 先頭	フレーム 長	MC 識別	CID	振分 完了	MC カウンタ	フレーム アドレス
0	4	0	0	0	0	0	0	0	0	0	0
1	13	0	0	0	0	0	1	2	0	0	0
2	3	0	0	0	0	0	0	0	0	0	0
3	11	0	0	0	0	0	1	0	0	0	5
4	2	0	0	0	0	0	0	0	0	0	0
5	7	1	0	0	1	3	1	2	1	2	0
6	0	2	1	3	1	4	0	0	0	0	0
7	1	0	0	0	0	0	1	2	0	0	0
8	12	0	0	0	0	0	0	0	0	0	0
9	XX	0	0	0	0	0	0	0	0	0	0
10	9	0	0	0	0	0	0	0	0	0	0
11	11	0	0	0	0	0	1	0	0	0	5
12	XX	0	0	0	0	0	0	0	0	0	0
13	8	12	0	0	1	3	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1

E: 次先頭アドレスイネーブル

アドレス管理 FIFO

MCキュー

12	8	13
----	---	----

フレーム

出力バッファキュー

3	2	4	0	6
---	---	---	---	---

MC フレーム

出力バッファキュー

11

MC

空きアドレスキュー

9	10
---	----

スタートエンドポイント

MCキュー

13	12
----	----

出力バッファキュー

6	3
---	---

出力バッファキュー

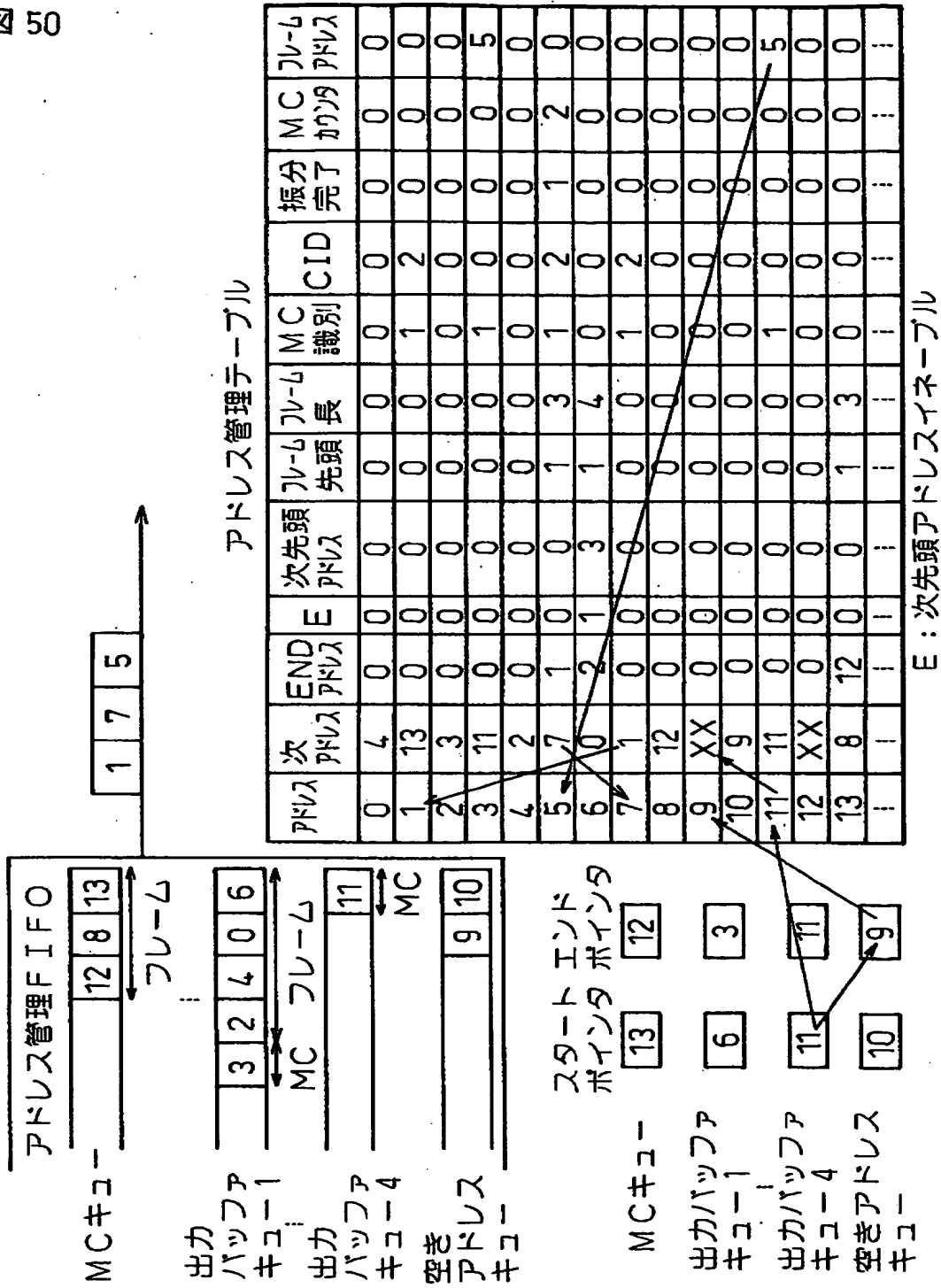
11	11
----	----

空きアドレスキュー

10	9
----	---

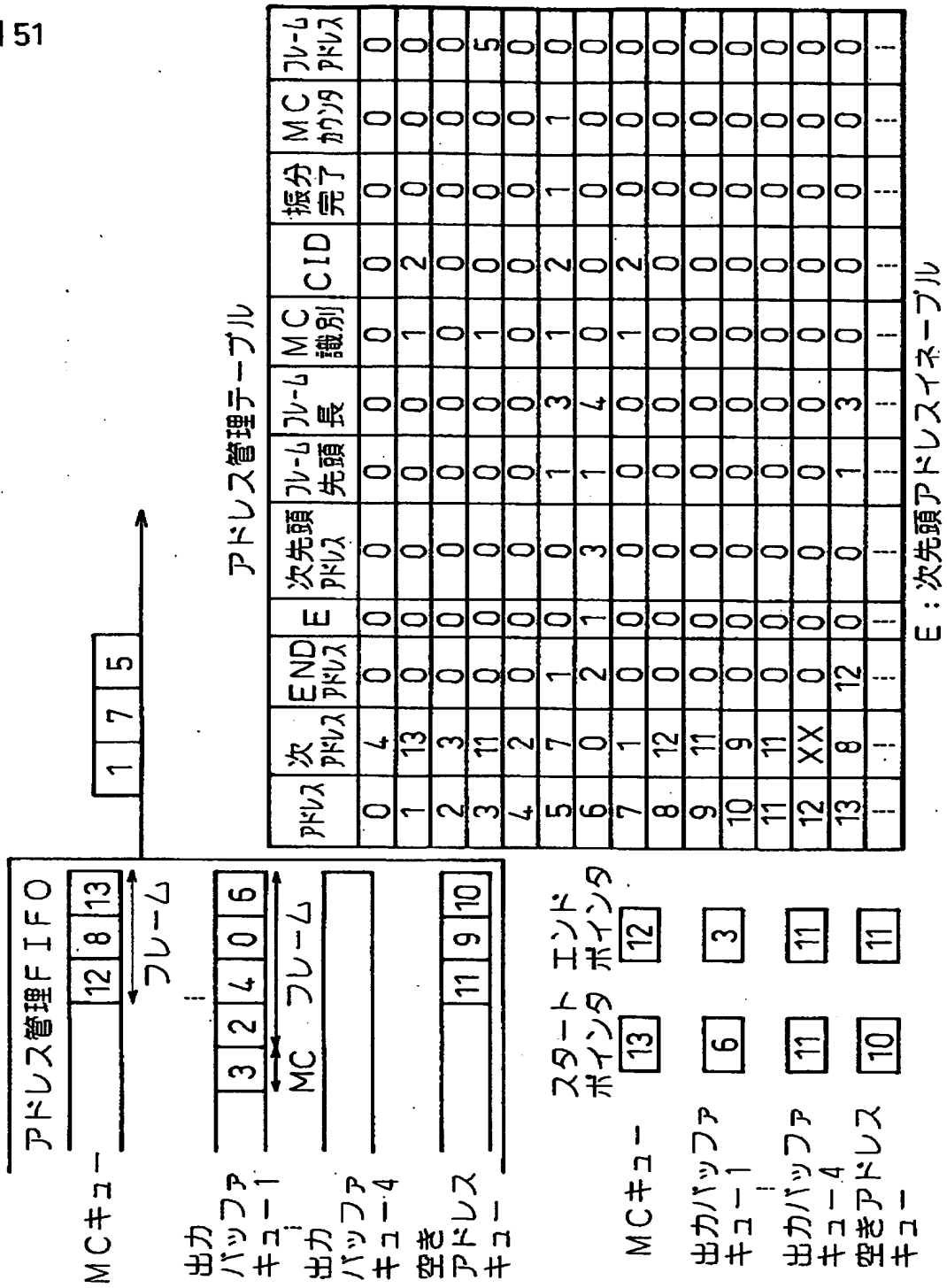
【図 50】

図 50



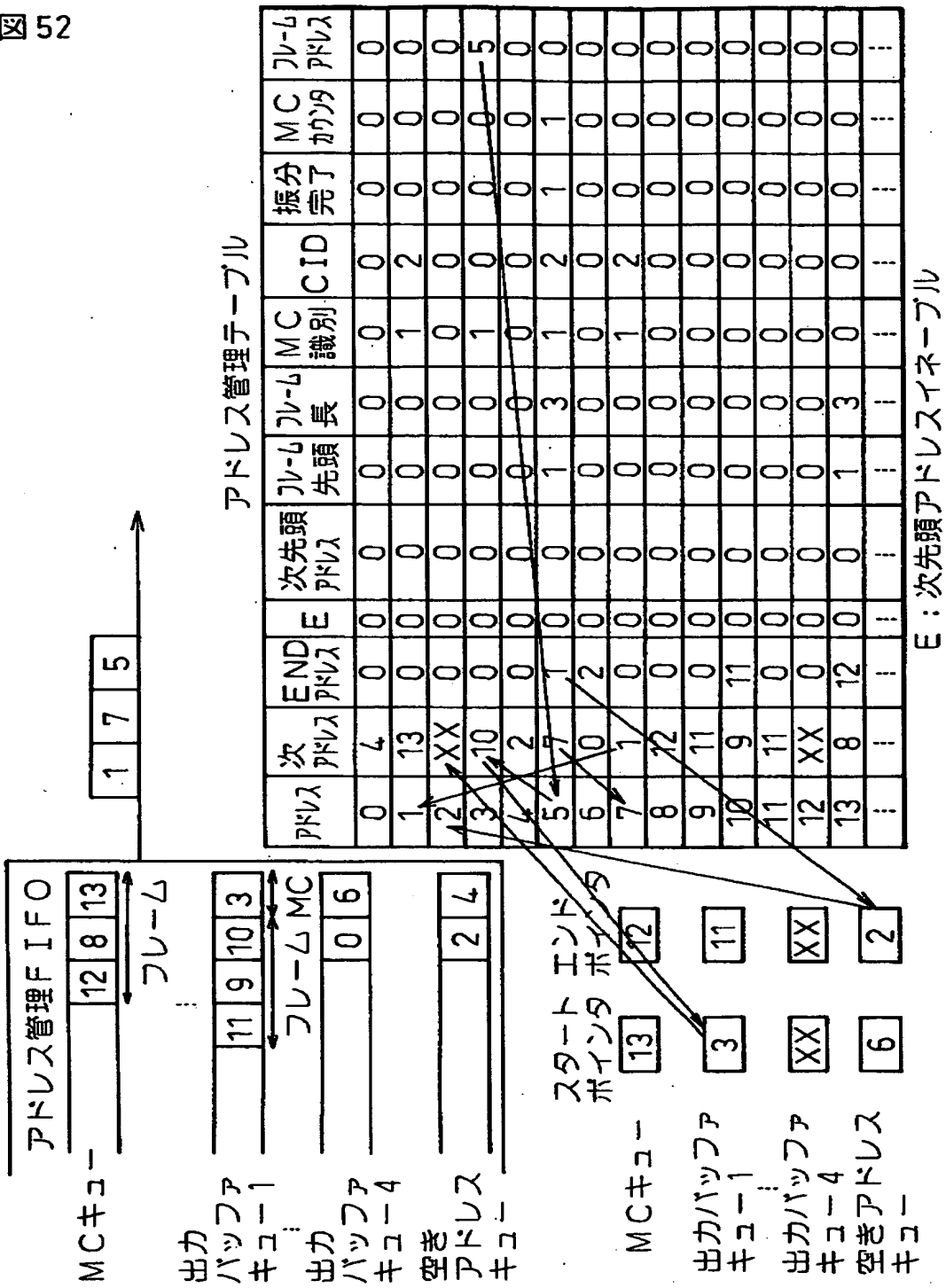
【図51】

図51



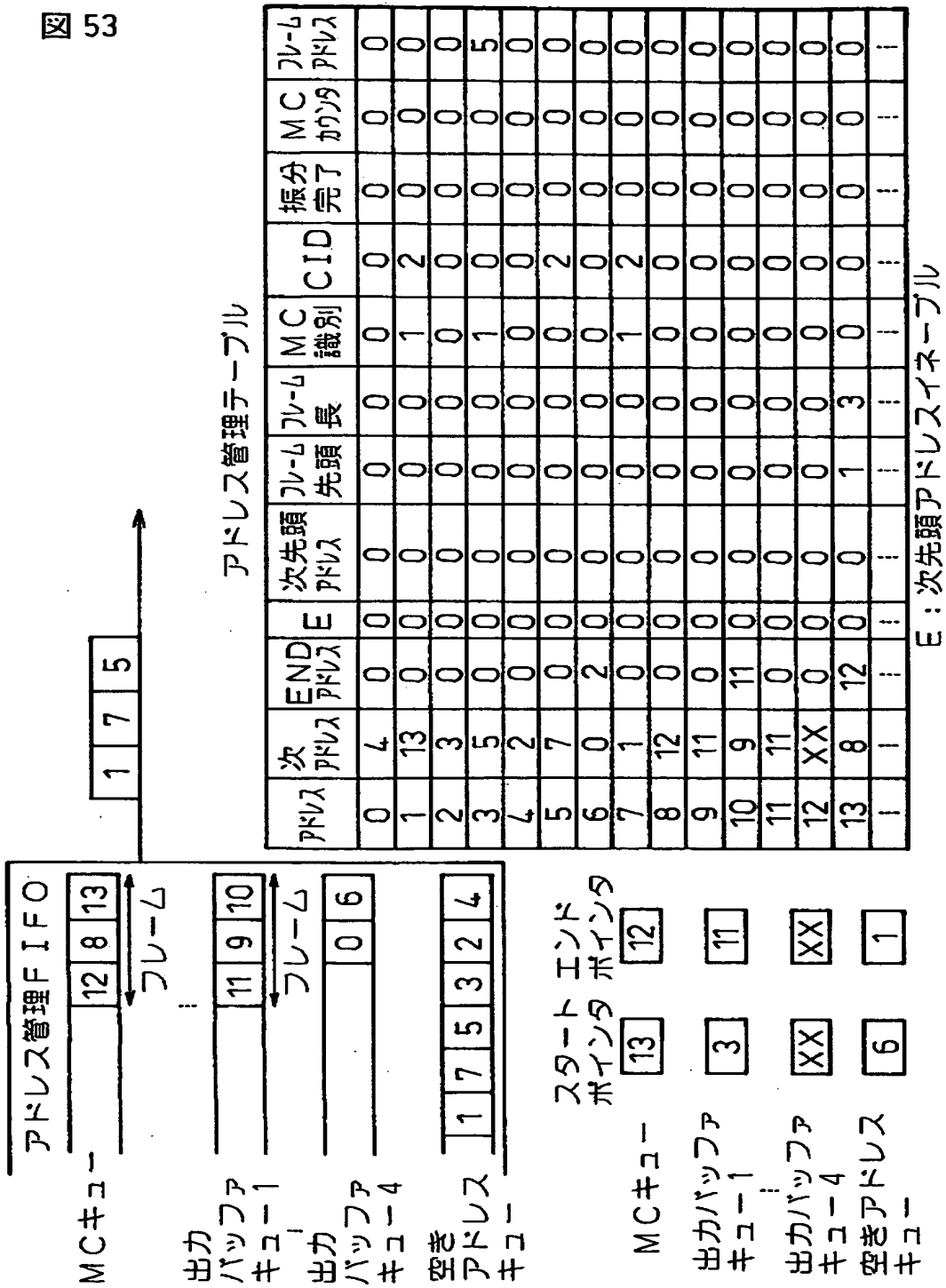
【図 52】

図 52



【図53】

図 53



【書類名】 要約書

【要約】

【課題】 可変長パケットに対して、高速かつ少ないハードウェア規模でQoS制御、廃棄制御およびマルチキャスト制御を実現する。

【解決手段】 パケット分割部10において、可変長パケットを固定長パケットに分割し、入力バッファ部12において、出力回線別さらにQoSクラス別にキューに格納する。この際に、多数のQoSクラスを、指定された帯域が保証される帯域保証クラスおよび余剰帯域が配分されるベストエフォートクラスの2種類のみマッピングすることにより、回線間スケジューラ14による入力側のスケジューリングを実現する。出力バッファ部18においてスイッチ部16でスイッチングされた固定長パケットから可変長パケットが組み立てられ、パケット長に依存したQoS制御が行なわれる。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [000005223]

1. 変更年月日	1996年 3月26日
[変更理由]	住所変更
住 所	神奈川県川崎市中原区上小田中4丁目1番1号
氏 名	富士通株式会社